ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

ALPEN-ADRIA UNIVERSITÄT KLAGENFURT | WIEN GRAZ

INHALTSREGULIERUNG AUF INTERNET-PLATTFORMEN

OPTIONEN FÜR VERANTWORTUNGSORIENTIERTE GOVERNANCE AUF NATIONALER EBENE

Forschungsbericht

Florian Saurwein

Charlotte Spencer-Smith

Institut für vergleichende Medien- und Kommunikationsforschung der
Österreichischen Akademie der Wissenschaften und der Alpen-Adria-Universität Klagenfurt

Wien, im September 2019

INSTITUT FÜR VERGLEICHENDE MEDIEN- UND KOMMUNIKATIONSFORSCHUNG

Offenlegung

Der vorliegende Forschungsbericht präsentiert die Ergebnisse einer Studie zum Thema Inhaltsregulierung auf Internet-Plattformen, die 2018/19 am Institut für vergleichende Medien- und Kommunikationsforschung der Österreichischen Akademie der Wissenschaften und der Alpen-Adria-Universität Klagenfurt entstand.

Die Studie wurde durch Sky finanziell unterstützt. Grundlagen für die Studie wurden im Rahmen des Projekts „Die Automatisierung des Sozialen: Algorithmische Selektion in Sozialen Online Netzwerken“ erarbeitet, das aus Mitteln des Jubiläumsfonds der Stadt Wien für die Österreichische Akademie der Wissenschaften unterstützt wird.

Die Ergebnisse der Studie entsprechen der Meinung der Autor/innen und widerspiegeln nicht zwangsläufig auch die Meinungen der unterstützenden Institutionen.

Kontakt und Rückfragen

Dr. Florian Saurwein
Institut für vergleichende Medien- und Kommunikationsforschung der
Österreichischen Akademie der Wissenschaften und der Alpen-Adria-Universität Klagenfurt
Postgasse 7 / 4 / 1, A-1010 Wien
florian.saurwein@oeaw.ac.at | +43 1 51581 3119

Inhaltsverzeichnis

OFFENLEGUNG 2
INHALTSVERZEICHNIS 3
VORWORT 4
KURZFASSUNG 5

HINTERGRUND UND PROBLEMSTELLUNG 8
ZIELSETZUNG UND FRAGESTELLUNGEN 10
VORGEHENSWEISE UND AUFBAU DER UNTERSUCHUNG 11

1. INHALTE AUF INTERNET-PLATTFORMEN: RISIKEN UND REGULIERUNGSBEDARF 12

2. GOVERNANCE-CHOICE: KONTEXTBEDINGUNGEN FÜR INHALTSREGULIERUNG 15
2.1 ANALYSERAHMEN 15
2.2 ÜBERPRÜFUNG DER KONTEXTFAKTOREN 16
2.3 FAZIT UND SCHLUSSFOLGERUNGEN ZU KONTEXTBEDINGUNGEN FÜR INHALTSREGULIERUNG 31

3. INHALTSREGULIERUNG AUF INTERNET-PLATTFORMEN: STATUS QUO IN EUROPA UND ÖSTERREICH 34
3.1 SELBSTORGANISATION: PRAXIS DER INHALTSREGULIERUNG DURCH DIE PLATTFORMBETREIBER 34
3.2 RECHTLICHE REGULIERUNG VON INTERNET-PLATTFORMEN UND INHALTEN 46
3.3 REGULIERUNG UND REFORMINITIATIVEN AUF EU-EBENE 51
3.4 SELBSTREGULIERUNG AUF EUROPÄISCHER EBENE 54
3.5 AUSGEWÄHLTE INITIATIVEN IN ANDEREN EUROPÄISCHEN LÄNDERN 56
3.6 FAZIT ZUM STATUS QUO DER AKTUELLEN REGULIERUNGSPRAXIS 59

4. GRUNDSÄTZE UND INSTITUTIONELLE KOMPONENTEN EINES VERANTWORTUNGSORIENTIERTEN GOVERNANCE-ARRANGEMENTS 63
4.1 RAHMENKONZEPT & LEITPRINZIP: VERANTWORTLICHKEIT (ACCOUNTABILITY) 63
4.1.1 GRUNDSÄTZE FÜR DIE INHALTSREGULIERUNG 64
4.1.2 GRUNDSÄTZE FÜR DAS GOVERNANCE-ARRANGEMENT 68
4.2 INSTITUTIONELLE KERNKOMPONENTEN (OPTIONEN, IDEENSAMMLUNG) 70
4.2.1 INSTITUTIONEN FÜR VERANTWORTUNGSORIENTIERTE KOMMUNIKATION 71
4.2.2 INSTITUTIONEN FÜR MELDE- UND BESCHWERDEMANAGEMENT 72
4.2.3 INSTITUTIONEN FÜR AUFSICHT UND KONTROLLE 78
4.3 FAZIT UND SCHLUSSFOLGERUNGEN ZU GRUNDSÄTZEN UND INSTITUTIONALISIERUNGSOPTIONEN 84

5. REFORMOPTIONEN FÜR ÖSTERREICH 86
5.1 RAHMENBEDINGUNGEN & -STRATEGIEN 86
5.2 INSTITUTIONELLE REFORMOPTIONEN 87
5.3 FAZIT ZU REFORMOPTIONEN FÜR ÖSTERREICH 96

LITERATURVERZEICHNIS 99

Vorwort

Die rasche Verbreitung von Internet-Plattformen wie Google, Facebook und YouTube, neue Anbieter im Medienmarkt wie Netflix und Amazon prime, neue und leistungsfähigere (mobile) konvergente Endgeräte und Infrastrukturen, die streaming von Bewegtbildern anytime und anywhere erlauben, und der damit verbundene Wandel der Nutzungsgewohnheiten und Refinanzierungsbedingungen stellt sogenannte „traditionelle" Medien ebenso vor Herausforderungen wie Medienpolitik und Medienforschung. Wie sollen „legacy media" auf die Konkurrenz durch Internet-Plattformen reagieren? Welcher Handlungsbedarf besteht für die Medienpolitik um Risiken für Vielfalt, Qualität und Verantwortung zu begegnen? Welche wissenschaftlichen Ansätze ermöglichen es uns, die Umbrüche im Mediensystem adäquat zu erfassen und Gefahren zu erkennen, um Politik und Unternehmen bestmöglich zu beraten und die Medienkompetenz in der Gesellschaft zu verbessern?

Das Institut für vergleichende Medien- und Kommunikationsforschung der Österreichischen Akademie der Wissenschaften und der Alpen-Adria-Universität Klagenfurt widmet sich der Erforschung des Medienwandels, den damit verbundenen Veränderungen für politische Kommunikation und öffentliche Meinungsbildung sowie den daraus resultierenden Herausforderungen für Governance und Verantwortlichkeit auf der Mikro-, Meso- und Makroebene des gesellschaftlichen Kommunikationssystems. Seine Forschungserkenntnisse sollen sowohl in den wissenschaftlichen Diskurs und die universitäre Lehre einfließen als auch als praxisrelevante Beiträge die öffentliche und politische Debatte bereichern.

Die vorliegende Studie trägt diesem Anspruch zur Verbindung von Forschung und Praxis in vielfacher Weise Rechnung. Internet-Plattformen sind von steigender Bedeutung für öffentliche Diskurse in der mediatisierten Gesellschaft und entsprechend steigt auch ihre Bedeutung als Forschungsgegenstand. Inhaltsregulierung oder „Content Moderation" auf Internet-Plattformen bewegt sich im brisanten Spannungsfeld von Freiheit und Verantwortung das auf zentrale normative Grundsätze verweist, die nicht immer in Einklang miteinander stehen und vielfach nach Interessenabwägung verlangen. Internet-Plattformen stehen in der Kritik dieser Anforderung bislang nicht adäquat zu begegnen. Die in der vorliegenden Studie vorgeschlagene Idee einer nationalen Ombudsstelle für die Berufung gegen Moderationsentscheidungen der Plattformen und für Streitbeilegung stellt eine praxisrelevante Empfehlung dar, um kontextsensible Interessenabwägungen zu ermöglichen. Ein adäquates institutionelles Design einer Ombudsstelle sollte deshalb eine breite Stakeholdereinbindung zum Interessenausgleich sicherstellen, und eine Spruchpraxis ermöglichen, die sowohl ethische und als auch rechtliche Normen der Inhaltsregulierung beachtet. Insgesamt geht es dabei nicht zuletzt um einen Beitrag zur Etablierung einer Verantwortungskultur in der Online Kommunikation, die in Zeiten von Desinformation, Skandalisierung und Hate Speech besonders wichtig erscheint.

Wien, im September 2019 Matthias Karmasin

Kurzfassung

Die Verbreitung des Internet hat die gesellschaftliche Informations- und Kommunikationsstruktur in nur wenigen Jahren revolutioniert. Die neuen kommunikativen Möglichkeiten fördern die freie Meinungsäußerung und die Beteiligung an öffentlichen Debatten. Die Offenheit des Internet führt aber auch zur Verbreitung von illegalen Inhalten wie Kindesmissbrauchsdarstellung und Terrorismus, zur Verletzung von Persönlichkeitsrechten und der Menschenwürde durch Beleidigung, Verleumdung, Hassrede und Cybermobbing, zu Urheberrechtsverletzungen und zur Verbreitung unerwünschter Werbung.

Speziell die großen, viel genutzten Internet-Plattformen wie Facebook, Twitter und YouTube stehen vor der Herausforderung adäquat mit illegalen und unerwünschten Inhalten umzugehen, die über ihre Dienste verbreitet werden. Dabei genießen die Plattformen eine Haftungsprivilegierung, indem sie nicht gegen rechtswidrige Inhalte vorgehen müssen, solange sie keine Kenntnis davon haben. Sie etablieren jedoch selbst Gemeinschaftsstandards und komplexe Systeme zur Inhaltsregulierung, um problematische Inhalte zu identifizieren, zu filtern, zu löschen, im Rahmen der Kuratierung weniger prominent darzustellen oder zu kennzeichnen (Inhaltsmoderation).

Durch die weltweit starke Nutzung der Plattformen spielt die Inhaltsmoderation eine immer größere Rolle für die Vermittlung von Information und für die Wahrnehmung der Welt. Plattformbetreibern werden u.a. Eingriffe in die Meinungsfreiheit und mangelnde Sensibilität für nationale Bedürfnisse vorgeworfen. Problematisiert wird zudem, dass private Akteure private Regeln etablieren und durchsetzen, damit jedoch öffentliche Kommunikation beeinflussen, die für das Funktionieren von demokratischen Gesellschaften bedeutend ist. Aufgrund der Kritik wird eine breite Debatte zum Thema Inhaltsregulierung auf Internet-Plattformen mit Blick auf Öffentlichkeit, Demokratie und Kontextsensibilität eingefordert, in der auch das Thema der *Verantwortlichkeit* eine immer bedeutendere Rolle spielt.

Die vorliegende Studie leistet einen Beitrag zur Debatte über Verantwortlichkeit für Inhaltsmoderation auf Internet-Plattformen. Sie analysiert die Potenziale und Grenzen einer reinen marktnahen Inhaltsregulierung durch individuelle Selbstorganisation der Unternehmen sowie kollektive Selbstregulierung auf Branchenebene und zeigt den Bedarf an öffentlicher Kontrolle der Inhaltsregulierung. Darauf aufbauend werden Grundsätze und Optionen für ein Governance-Arrangement identifiziert, das auf nationaler Ebene ansetzt und dazu beiträgt, die Verantwortlichkeit globaler Plattformen gegenüber jenen nationalen Öffentlichkeiten zu verbessern, die sie durch ihre Dienste und Moderationsentscheidungen mitgestalten.

Die zentralen Ergebnisse der Analysen lassen sich wie folgt zusammenfassen:

Die theoretische Analyse der Bedingungen für Inhaltsregulierung legt ein Regulierungsarrangement nahe, das in der operativen Praxis von den Plattformen selbst betrieben wird und gleichzeitig unter einer öffentlichen Kontrolle steht. Für Selbstorganisation und Selbstregulierung spricht vor allem, dass staatliche Eingriffe in Medieninhalte grundrechtsproblematisch

sind und deshalb Selbstorganisation und Selbstregulierung zu bevorzugen sind. Für Selbstregulierung sprechen zudem die immense Zahl an Inhalten im digitalen Umfeld, die nur von den Plattformbetreibern selbst moderiert werden kann, der indirekte Nutzen einer adäquaten Inhaltsregulierung für die Industrie (Image, Nutzererlebnis), die Reputationssensitivität der stark exponierten Plattformunternehmen, und die oligopolistische Marktstruktur, mit wenigen dominanten Playern, durch die Selbstregulierung erleichtert wird.

Gleichzeitig verdeutlicht die Analyse, dass Inhaltsmoderation nicht ausschließlich der Industrie überlassen werden soll. Internet-Plattformen agieren als Kuratoren und Regulatoren von Inhalten mit hoher Eingriffsintensität durch automatisierte Inhaltskontrolle im grundrechtssensiblen und demokratierelevanten Bereich gesellschaftlicher Kommunikation. Dies erfordert Verantwortlichkeit und externe Kontrolle, um Willkür zu verhindern, die Rechte der Betroffenen zu wahren, faire Verfahren zu ermöglichen und um sicher zu stellen, dass Inhaltsregulierung transparent und rechtmäßig erfolgt.

Internet-Plattformen verfügen zudem über eine hohe Marktmacht, einen starken Einfluss auf gesellschaftliche Kommunikation und über immer mehr politische Macht. Daher besteht ein Bedarf an einer stärkeren Verteilung von Verantwortlichkeit und an „checks and balances" zur Machtkontrolle im System zur Inhaltsregulierung, in dem bislang die Plattformen selbst oft gleichzeitig als Legislative, Exekutive und Judikative agieren.

Schließlich sprechen die ökonomischen und organisatorischen Interessen der Internet-Plattformen tendenziell für möglichst automatisierte Inhaltsmoderation, möglichst internationale Standards und globale institutionelle Arrangements zur Inhaltsregulierung. Solche globalen, automatisierten Lösungen gehen jedoch auf Kosten von regionalen Interessen und Kontextsensitivität, und sie erschweren umsichtigen Abwägungen im Einzelfall. Eine zentrale Herausforderung für institutionelle Arrangements besteht darin, kontextsensible Ansätze für Inhaltsregulierung zu etablieren, ohne die Funktionsfähigkeit der Plattformen zu gefährden.

Abschließend wird daher gezeigt, welche institutionellen Kernkomponenten eine nationale Rückkoppelung der Inhaltsregulierung ermöglichen und welche Reformoptionen sich in Österreich anbieten, um verantwortungsorientierte Kommunikation, Melde- und Beschwerdemanagement sowie Aufsicht und Kontrolle im nationalen Rahmen zu fördern.

Die Studie kommt zu folgenden Empfehlungen:

Die Plattformen sollten *nationale Sprecher/innen* einsetzen, die dazu bevollmächtigt sind, Auskunft zu erteilen und auf Fragen zu antworten, damit die kommunikative Verantwortung gegenüber nationalen Öffentlichkeiten besser wahrgenommen wird.

Im Melde- und Beschwerdemanagement kann geteilte Verantwortung in Phasen mehrstufiger Prozesse in „kaskadischen Systemen" etabliert werden, die plattforminterne und -externe Meldestellen, plattforminterne Moderationsverfahren sowie plattformexterne Berufungsmöglichkeiten umfassen.

So bilden die bestehenden plattformexternen *Meldestellen* für problematische Inhalte auf nationaler Ebene wichtige Seismographen und Knotenpunkte zwischen nationalen Öffentlichkeiten und den Plattformen. Der Bestand und die Entwicklung der Meldestellen, die Zusammenarbeit mit den Plattformen (z.B. Trusted Flagger-Programme), ihre Bekanntheit und ihre Beratungsangebote sind von öffentlichem Interesse und sollen entsprechend öffentlich gefördert werden.

Die Plattformen selbst sollen Moderationssysteme ausbauen und dabei auch *Moderatorinnen und Moderatoren mit nationalem Kontextwissen* beschäftigen, um Moderationsentscheidungen zu ermöglichen, die regionalen und nationalen Gegebenheiten adäquat Rechnung tragen.

Für die Berufung gegen Moderationsentscheidungen und für Streitbeilegung bietet sich die Einrichtung einer *nationalen Ombuds- und Schlichtungsstelle* an, die in Kooperation mit den Plattformen agiert. Eine Ombudsstelle soll als Forum mit Perspektivenvielfalt durch breite Stakeholdereinbindung gestaltet sein. Sie soll strittige Fälle prüfen und gegenüber Plattformen Empfehlungen aussprechen, zu denen sie aufgrund einer umsichtigen Interessenabwägung kommt. So trägt sie zu einer Verteilung von Verantwortlichkeit bei und kann einige der Nachteile des Rechtsweges (Prozesskostenrisiko) und plattforminterner Verfahren (mangelndes Kontextverständnis) kompensieren.

Darüber hinaus sind Ombudsstellen aufgrund ihrer Problemnähe geeignet *Aufsichts- und Monitoringfunktionen* zu übernehmen, Entwicklungen und Herausforderungen der Plattformkommunikation und Inhaltsmoderation systematisch zu beobachten und sich im internationalen Verbund an Audits und Evaluierungen der Inhaltsregulierung durch Plattformen zu beteiligen.

Zur Ausübung ihrer Berufungs- und Aufsichtsfunktionen müssen einer Ombudsstelle relevante Informationen und Daten von den Internet-Plattformen zur Verfügung gestellt werden. Wenn dies im Weg einer freiwilligen Kooperation mit den Plattformen scheitert, sollten die Regulierungsbehörden zusätzliche Maßnahmen in Erwägung ziehen, um die Kooperation der Plattformen sicher zu stellen. Beispielsweise durch Anreize wie eine Staffelung der Werbe- bzw. einer möglichen Digitalsteuer in Abhängigkeit von der Kooperationsbereitschaft, oder durch die Einrichtung von Ko-Regulierung auf Basis einer gesetzlichen Grundlage, mit der Plattformunternehmen zu Information und Kooperation verpflichtet werden und die geeignete Sanktionsmöglichkeiten vorsieht.

Hintergrund und Problemstellung

Die rasante Verbreitung des Internet hat die gesellschaftliche Informations- und Kommunikationsstruktur in nur wenigen Jahren revolutioniert. Internetnutzern steht eine Vielzahl an Informationen und Informationsquellen und zur Verfügung, die ihre Inhalte überwiegend kostenlos anbieten. Die traditionellen Geschäftsmodelle im Mediensektor wie Abonnement- und Werbefinanzierung geraten zunehmend unter Druck. Für Anbieter und Sprecher war es noch nie so einfach Zugang zur Öffentlichkeit zu erlangen, Informationen zu verbreiten und Meinungen kundzutun. Auch wenn nicht alle Stimmen die gleichen Chancen haben, auch gehört zu werden, sind mit diesen neuen Möglichkeiten zur Artikulation, Verbreitung und Nutzung von Information viele positive Entwicklungen verbunden. Sie fördern die freie Meinungsäußerung, verbessern die Offenheit der Kommunikationsinfrastruktur und verbreitern die Möglichkeit zur Beteiligung an öffentlichen Debatten. Die Vorteile des offenen Internet sind jedoch auch von etlichen Barrieren und Risiken begleitet. So führt die zunehmende Vielfalt an Informationen zu Orientierungsschwierigkeiten für Nutzer, und der Vielzahl an Informationen im Netz steht die Knappheit der Aufmerksamkeit auf Rezipientenseite gegenüber. Zudem führt die Offenheit des Internet auch zur Verbreitung illegaler und gefährlicher Inhalte wie Kindesmissbrauchsdarstellungen, Hassbotschaften, Belästigung, Desinformation und Extremismus.

Besonders den großen Internet-Plattformen wie Facebook, Twitter, Google und Amazon ist es gelungen mit innovativen technologischen Lösungen Orientierungsprobleme zu reduzieren und dazu beizutragen, den Information Overload zu bewältigen. Mit Hilfe von Algorithmen werden Inhalte entsprechend der prognostizierten Nutzerinteressen kuratiert. Sprecher und Publikum, Angebot und Nachfrage können so gezielt miteinander verbunden werden. Die Automatisierung, Skalierbarkeit und die daraus resultierende Effektivität der Vermittlungsleistung machen einen großen Teil des Erfolges der Internet-Plattformen aus. Aufgrund ihrer Vermittlungsfunktion werden die Plattformen auch als „Internet-Intermediäre“ bezeichnet.

Dieser Fokus auf ihre vermittelnde, intermediäre Funktion führt geradewegs zur Frage der Verantwortung für Inhalte, die von Dritten produziert und über ihre Plattformen vermittelt werden. Plattformbetreiber sehen sich primär als Technologiedienstleister und lehnen eine unmittelbare Verantwortlichkeit für Inhalte Dritter ab. Etliche geltende rechtliche Regulierungen stützen diese Sichtweise und bieten eine substanzielle Haftungsprivilegierung für Plattformbetreiber. Mit neueren Regulierungsinitiativen wie dem Netzwerkdurchsetzungsgesetz (NetzDG) in Deutschland und der Reform im EU-Urheberrecht scheint dieses Prinzip jedoch langsam zu erodieren.

Zudem etablieren auch die Plattformbetreiber selbst Community-Standards um „problematische“ Inhalte auf ihren Plattformen zu verhindern und positive Nutzererlebnisse zu fördern. Dafür implementieren die Internet-Plattformen komplexe Systeme zur Kontrolle von Inhalten, um problematische Beiträge zu identifizieren, zu filtern, zu löschen, zu kennzeichnen oder im Rahmen der Kuratierung weniger prominent darzustellen (content moderation). Die Moderation von Inhalten auf Internet-Plattformen ist eine arbeitsintensive und anspruchsvolle Aufgabe, die mittels einer Kombination aus automatisierten Verfahren und menschlichen Kontrollen erfolgt. Die Inhaltsmoderation gilt mittlerweile als konstitutives Merkmal der Intermediäre und unterscheidet die Internet-Plattformen auch vom offenen Internet (Gillespie 2018). Durch die weltweit starke Nutzung der Plattformen spielen Moderation und Kuratierung zudem eine immer größere Rolle für die Vermittlung von Information und damit letztlich auch für die Wahrnehmung der Welt. Eingriffe durch Kuratierung und Moderation wirken regulierend auf gesellschaftliche Information und Kommunikation. Öffentlichkeit und Kultur werden durch Inhaltsregulierung auf Internet-Plattformen mitgeprägt. Kuratierung und Moderation sind einflussreich und folgenhaft. Vielfach wird kritisiert, dass die damit verbundene Macht unkontrolliert in den Händen der Plattformbetreiber liegt und dass die mit Moderation und Kuratierung verbundenen Maßnahmen zu wenig transparent ausgeübt werden.

Plattformbetreibern werden Eingriffe in die Meinungsfreiheit und mitunter auch Zensur vorgeworfen. Kritisiert wird, dass die Unternehmensstandards für Inhaltsregulierung nicht vollständig veröffentlicht und nachvollziehbar sind und dass vielfach unklar bleibt, aufgrund welcher Kriterien „akzeptable“ von „inakzeptablen“ Inhalten unterschieden werden. Egal ob Inhalte gelöscht werden oder sichtbar bleiben, für Beschwerdeführer, deren Reklamationen zurückgewiesen werden, ist der Verweis der Plattformbetreiber auf „Community-Standards“ meistens wenig befriedigend – und der Rechtsweg als Alternative zu aussichtlos, zu langwierig und zu kostspielig. Problematisiert wird zudem, dass private Akteure private Regeln etablieren und mit privaten Mitteln und Methoden durchsetzen – und damit jedoch gleichzeitig öffentliche Information und Kommunikation prägen und beeinflussen, die für das Funktionieren von demokratischen Gesellschaften bedeutend sind. Schließlich wird auch kritisiert, dass Inhaltsregulierung auf Plattformen sehr zentral organisiert ist und zu wenig auf nationale und regionale Kontexte Rücksicht nimmt.

Aufgrund dieser Kritik wird mittlerweile eine breite Debatte zum Thema Content Moderation mit Blick auf Öffentlichkeit, Demokratie und Kontextsensibilität eingefordert, in der auch das Thema Verantwortung eine immer bedeutendere Rolle spielt. Dabei soll diese Debatte über die bislang dominierende Frage der rechtlichen Verantwortlichkeit im Sinne einer „Haftung“ für Inhalte hinausgehen und die Frage der Verantwortung in einem breiteren Verständnis adressieren. Dafür bietet sich das Konzept der „Accountability“ an. Accountability kann als Mechanismus verstanden werden, der über die organisatorische Festlegung von Zuständigkeits- und Verantwortungsbereichen und über rechtliche Verursachungs- und Haftungsfragen hinausgeht und eine starke kommunikative Komponente beinhaltet, die sich mit Rechtfertigung, Rede-und-Antwort-Stehen oder „sich erklären“ gut beschreiben lässt. Bovens (2007) definiert

Accountability als eine Beziehung zwischen einem Akteur und einem Forum, in welcher der Akteur verpflichtet ist, sein Verhalten zu erklären und zu rechtfertigen, und in der das Forum Fragen stellen und Urteile fällen kann und der Akteur Konsequenzen zu tragen hat.

Für die Praxis stellt sich die Frage wie Verantwortlichkeit in diesem breiten Verständnis für das Regulierungsfeld der Inhaltsmoderation auf Internet-Plattformen adäquat implementiert werden kann; welche institutionellen Governance-Arrangements – bestehend aus Normen und Organisationen – sich dafür anbieten; welche Rolle die privaten Plattformbetreiber, der Staat, Nutzer und unabhängige Dritte dabei spielen sollen; und wie Aufgaben und Kontrolle zwischen privaten und staatlichen Akteuren verteilt werden können.

Zielsetzung und Fragestellungen

Ziel der Studie ist es, Prinzipien, Komponenten und Reformvorschläge für ein Governance-Arrangement zu identifizieren, das sich dafür eignet, die Verantwortlichkeit für die Moderation von Internet-Inhalten durch Plattformbetreiber in Zukunft stärker national in Österreich zu verankern. Dabei geht es nicht primär um alle Details einer rechtlichen Ausgestaltung der Verantwortung für Internet-Inhalte, sondern um die Benennung von Grundsätzen, institutionellen Kernkomponenten und Reformüberlegungen für ein verantwortungsorientiertes Governance-Arrangement, das auf nationaler Ebene umgesetzt werden könnte. Im Detail soll die Studie folgende Fragen beantworten:

- Warum besteht ein Bedarf an einer Regulierung von Inhalten im Internet durch Kontrolle, Kuratierung und Filterung?
- Wo liegen die Stärken und Schwächen der Regulierung von Internet-Inhalten durch Plattformbetreiber und durch staatliche Akteure und welche Aufgabenverteilung zwischen Markt und Staat bietet sich aus theoretischer Perspektive an?
- Wie ist die Regulierung von Internet-Inhalten in Österreich und Europa aktuell (durch Recht und industrielle Selbstregulierung) institutionalisiert und welche Schwächen weist die aktuelle Regulierungspraxis auf?
- Welche Grundsätze und institutionelle Kernkomponenten gelten als zentral für ein Governance-Arrangement, das die Verantwortlichkeit von Internet-Intermediären gegenüber Nutzern, Öffentlichkeit und Politik in den Mittelpunkt rückt?
- Welche institutionellen Reformen (Normen, Organisationen) wären notwendig, um ein solches verantwortungsorientiertes Governance-Arrangement in Österreich zu implementieren?

Vorgehensweise und Aufbau der Untersuchung

Die Studie basiert auf einer systematischen Literaturrecherche an der Schnittstelle der Themen Internet-Plattformen, Informationsintermediäre, illegale und gefährliche Inhalte, Inhaltskontrolle und Content Moderation. Die Beantwortung der Fragen erfolgt mit Hilfe von Sekundäranalysen und Zusammenfassungen der Kernergebnisse bestehender Literatur zum Thema Inhaltsregulierung und Verantwortlichkeit. Ergänzt wird die Sekundäranalyse durch eine theoretische Analyse der Aufgabenverteilung zwischen Markt und Staat für die Moderation von Internet-Inhalten mit Hilfe des Governance-Choice-Ansatzes sowie sozialwissenschaftliche Analysen zum Status quo und den Reformoptionen für die Governance der Inhaltsregulierung in Österreich. Zur Abschätzung und Validierung der praktischen Umsetzbarkeit von tentativen Überlegungen zu Reformoptionen wurden Interviews mit Expertinnen und Experten in Österreich durchgeführt.[1]

Im Detail ist die Untersuchung wie folgt aufgebaut:

Abschnitt 1 bietet einen kurzen Überblick zu *Risiken und Problembereichen der Verbreitung von Inhalten* über Internet-Plattformen. Dazu zählen etwa Kindesmissbrauchsdarstellungen, Hassbotschaften, Missbrauch und Belästigung, Desinformation und Extremismus. Die Risiken und Problembereiche bilden gleichzeitig die Begründung für den Bedarf einer Regulierung von Internet-Inhalten (Warum soll bzw. muss reguliert werden?).

In Abschnitt 2 erfolgt ein Assessment der Regulierungsoptionen am Kontinuum zwischen Markt und Staat. Mit Hilfe des *Governance-Choice-Ansatzes* werden die Kontextbedingungen für die Regulierung von Inhalten auf Internet-Plattformen dahingehend überprüft, inwieweit sie sich für eine industrielle Selbstregulierung eignen oder ob ein Bedarf an öffentlicher Kontrolle und staatlicher Involvierung in der Inhaltsregulierung besteht. Die theoretische Analyse dieser Kontextfaktoren zeigt, welches Potenzial und welche Grenzen die Selbstregulierung hat und welche Kontextbedingungen für eine öffentliche Kontrolle sprechen. Im Ergebnis liefert Abschnitt 2 Schlussfolgerungen in Form von Empfehlungen für die Verteilung von Regulierungsaufgaben zwischen Markt und Staat aus theoretischer Perspektive.

Abschnitt 3 bietet eine Beschreibung des Status quo der aktuellen Regulierungspraxis (Wie wird aktuell reguliert?). Diese umfasst zuerst die Maßnahmen zur Inhaltsregulierung, die von Internet-Plattformen gesetzt werden (Selbstorganisation, Selbstregulierung). Im Anschluss erfolgt ein Überblick zu den Eckpfeilern des aktuell geltenden Regulierungsrahmens in Österreich und der Europäischen Union. Ergänzt wird diese Erhebung durch eine Analyse zu Schwächen im aktuellen Regulierungsarrangement anhand ausgewählter Kritikpunkte, die

[1] Die Gespräche mit nationalen Expertinnen und Experten fanden im Zeitraum Februar bis Mai 2019 statt. Interviewpartnerinnen und -partner waren Maximilian Schubert (ISPA), Bernhard Jungwirth und Karl Gladt (ÖIAT, Internet Ombudsmann), Caroline Kerschbaumer und Hans Dieter Schindlauer (ZARA), Ben Wagner (WU Wien), Ingrid Brodnig, Klaudia Zotzmann-Koch und Maria Windhager.

im Fachdiskurs geäußert werden (Welche Defizite zeigt das aktuelle Regulierungssystem?). Dazu zählt z.B. die Kritik an fehlender Kontextsensibilität der Verfahren der Inhaltsregulierung durch private Akteure. Im Ergebnis liefert Abschnitt 3 einen Überblick zum Stand und zur Kritik der aktuellen Regulierungslage als Begründung für den anstehenden Reformbedarf.

Abschnitt 4 bietet einen Überblick zu Grundsätzen/Prinzipien und Kernelementen für die Ausgestaltung eines institutionellen Governance-Arrangements zur Verankerung von Verantwortlichkeit für die Moderation von Internet-Inhalten durch Plattformbetreiber. Mit Hilfe von Literaturanalysen werden Vorschläge für Prinzipien und ihre institutionelle Umsetzung identifiziert, die in der Diskussion über Inhaltsmoderation gemacht werden. Dazu zählen beispielsweise Transparenz, Verhältnismäßigkeit, Ombudsstellen und Beschwerdeverfahren sowie unabhängige Aufsichtsstellen zur externen Kontrolle. Im Ergebnis liefert Abschnitt 4 einen Überblick zu Prinzipien und Optionen zur Umsetzung eines verantwortungsorientierten Governance-Arrangements.

In Abschnitt 5 wird abschließend erörtert, welche Reformschritte notwendig wären, um ausgewählte Komponenten eines verantwortungsorientierten Governance-Arrangements in Österreich zu implementieren. Dabei geht es nicht um Details einer rechtlichen Ausgestaltung von Haftung, Rechtsdurchsetzung und Rechtsschutz für die Regulierung von Internet-Inhalten. Die Analyse fokussiert auf die Identifikation von geeigneten institutionellen Komponenten eines verantwortungsorientierten Governance-Arrangements auf nationaler Ebene in Österreich und umfasst die Identifikation des Handlungsbedarfs für die Plattformbetreiber (z.B. Etablierung von Auskunftspersonen), Vorschläge für die institutionelle Weiterentwicklung des Melde- und Beschwerdemanagements (z.B. Ombudsstellen) und Überlegungen zur Etablierung von Aufsichtsfunktionen im nationalen Rahmen.

1. Inhalte auf Internet-Plattformen: Risiken und Regulierungsbedarf

Die Verbreitung des Internet hat die gesellschaftliche Informations- und Kommunikationsstruktur in nur wenigen Jahren revolutioniert. Für Sprecher war es noch nie so einfach Zugang zur Öffentlichkeit zu erlangen, Informationen zu verbreiten und Meinungen kundzutun. Die neuen kommunikativen Möglichkeiten fördern die freie Meinungsäußerung und die Möglichkeit zur Beteiligung an öffentlichen Debatten. Die Vorteile des offenen Internet sind jedoch auch von Barrieren und Risiken begleitet. Zwar ist der überwältigende Teil der Inhalte im Internet harmlos. Die Offenheit des Internet, die niedrigen Zugangsbarrieren und die verbesserten Kommunikationsmöglichkeiten führen jedoch auch zur Verbreitung von illegalen und unerwünschten Inhalten, die einen *Bedarf* an Inhaltskontrolle und -regulierung begründen.

Zu solchen problematischen Inhalten zählen v.a. illegale Inhalte wie beispielsweise Kindesmissbrauchsdarstellungen, harte Pornographie, Terrorismus sowie speziell in Österreich und Deutschland die Holocaustleugnung und nationalsozialistische Wiederbetätigung. Des Wei-

teren umfassen unerlaubte Inhalte auch Verletzungen von Persönlichkeitsrechten, der Menschenwürde und der Privatsphäre zum Beispiel durch Beleidigung, Verleumdung, Verhetzung, Cybermobbing, Kreditschädigung, oder unerwünschte Werbung (Spam). Zudem ist die Verbreitung von Inhalten untersagt, mit denen Urheberrechte verletzt werden, etwa durch die Verletzung von Markenrechten, Rechten an geistigem Eigentum, oder Rechten am eigenen Bild. Aus gesellschaftspolitischer Perspektive handelt es sich dabei um Inhalte, die unerwünscht oder verboten sind, weil sie beispielsweise zu Gewalt auffordern, sich gegen Menschen richten oder Persönlichkeits- sowie Urheberrechte verletzen. Aus ökonomischer Perspektive handelt es sich um sogenannte *de-meritorische Güter*, die unter reinen Marktbedingungen, ohne regulatorische Eingriffe, in einem aus gesellschaftspolitischer (paternalistischer) Perspektive zu hohen Ausmaß produziert, verbreitet und konsumiert werden. Zum Schutz der betroffenen Individuen und wichtiger gesellschaftlicher Interessen wird die Produktion und Verbreitung solcher Inhalte untersagt oder beschränkt und rechtlich durchgesetzt.

Die Regulierung der Inhalte auf Internet-Plattformen umfasst in der Regel solche gesetzeswidrigen Inhalte, die gelöscht werden müssen, sofern die Plattformbetreiber davon Kenntnis erlangen. Die Inhaltsregulierung geht jedoch mitunter auch deutlich über die rechtlich vorgesehenen Beschränkungen hinaus. Die Internet-Plattformen entwickeln eigene Regeln (Community-Standards), die Verhaltensformen und Inhalte definieren, die auf den jeweiligen Plattformen unerwünscht sind. Dazu zählen beispielsweise legale, aber für bestimmte Nutzergruppen gefährliche oder aus anderen Gründen unerwünschte Inhalte wie Pornographie oder zuletzt Fake-Accounts und Desinformation. Beispielhaft dafür sind die Regeln von Facebook zur Vermeidung von Nacktheit.

Nach jahrelanger Kritik an der Intransparenz der Inhaltsregulierung wurden von einigen Internet-Plattformen 2018 erstmals Daten zur Inhaltsregulierung auf Basis der Community-Standards in einem Transparency Report dokumentiert.[2] Die publizierten Informationen verweisen u.a. auf den Umfang problematischer Inhalte, gegen die seitens der Plattformen verschiedene Maßnahmen gesetzt werden. So erfolgen aktuellen Daten (Herbst 2018) zufolge pro Quartal alleine bei Facebook über zwei Milliarden Maßnahmen zur Inhaltsregulierung in verschiedenen Problemkategorien wie:

- Nacktheit von Erwachsenen und Pornographie (20 - 30 Mio. Maßnahmen),
- Bullying und Belästigung (2 Mio.),
- Nacktheit von Kindern sowie Kindesmissbrauchsdarstellungen (9 Mio.),
- Fake-Accounts (600 - 800 Mio.),
- Hassrede (1,5 - 3 Mio.),
- Spam (800 Mio. - 1,5 Mrd.),
- Terroristische Propaganda (1 - 9 Mio.),
- Gewaltdarstellung (1 - 15 Mio.)
- Urheberechtsverletzungen (1,5 Mio.).

[2] Z.B. von Facebook: https://transparency.facebook.com/community-standards-enforcement

Maßnahmen gegen diese Arten von Inhalten werden von den Plattformbetreibern gesetzt, um rechtliche Standards einzuhalten und um Nutzern ein sicheres und angenehmes Nutzungserlebnis zu bieten. Die Inhaltsregulierungen durch Löschung, Kuratierung und andere Formen von Moderation gelten als konstitutive Merkmale der Internet-Plattformen (Gillespie 2018) und unterscheiden die Plattformen vom offenen Internet.

Dabei zeigt sich, dass sich die öffentliche Besorgnis in unterschiedlichen Ländern auch auf unterschiedliche Arten problematischer Inhalte fokussiert. In den USA und im Vereinigten Königreich[3] werden vor allem die Desinformationskampagnen rund um den Fall „Cambridge Analytica" und ihr möglicher Einfluss auf die vergangenen Wahlen und Referenden problematisiert. Hoch war auch die Sorge vor Desinformation und die Intransparenz von politischer Werbung im Vorfeld der Wahlen zum Europäischen Parlament 2019. In Österreich wiederum erhält das Thema Hasspostings besonders starke Aufmerksamkeit wobei vor allem seit der Migrationswelle 2015/16 häufiger strafrechtlich relevante Meinungsäußerungen wahrgenommen werden (Berka & Trappel 2019, 54). So hat beispielsweise der Verein für Zivilcourage und Anti-Rassismus-Arbeit (ZARA) für 2018 insgesamt 1.920 rassistische Vorfälle dokumentiert von denen 60,6% auf Hasspostings im Internet entfielen (ZARA 2019, 12 u. 43f.), und vorwiegend von sozialen Netzwerken (Facebook, Twitter, Google+, Xing etc.) stammen (ZARA 2018, 21). Rund zwei Drittel der österreichischen Internet-Nutzer haben schon Erfahrung mit Hasspostings gemacht (ISA 2017, 54). Zudem machte etwas mehr als die Hälfte der österreichischen Internetnutzer „sehr oft" oder „manchmal" Erfahrung mit Falschmeldungen (ISA 2017), die in den vergangenen Jahren verstärkt in Vorwahlzeiten und zu politisch polarisierenden Themen auftraten. Bei den Urhebern von Falschnachrichten handle es laut Einschätzung österreichsicher Expertinnen und Experten vorwiegend um Einzelpersonen, während der Einfluss organisierter, bezahlter Akteure zur Beeinflussung der Meinungsfreiheit bisher noch als kaum nennenswert beschrieben wird. Das Ausmaß, in dem Social Bots in Österreich eingesetzt werden, um vorrangig auf sozialen Netzwerken Einfluss auf die Meinungsbildung zu nehmen, wird von Expertinnen und Experten als wenig bekannt und erforscht bezeichnet (Berka & Trappel 2019, 66).[4]

Von der Verbreitung von Information werden allgemein positive *Informationsexternalitäten* erwartet. Menschen sollen sich über politische, soziale, wirtschaftliche und kulturelle Angelegenheiten informieren und daran teilnehmen. Meinungsfreiheit ist zentral. Durch die weltweit starke Nutzung spielen Plattformen dabei eine immer wichtigere Rolle – und mit ihnen auch die Moderation und Kuratierung von Inhalten. Inhalte und Inhaltsregulierung auf den großen, dominanten Social-Media-Plattformen sind von hohem öffentlichen Interesse und gesellschaftlicher Bedeutung. Sie gelten mittlerweile als wichtiger Teil der gesellschaftlichen und

[3] Siehe u.a. House of Commons (UK), Digital, Cultural, Media and Sport Committee (2019).

[4] Siehe dazu die Einschätzungen von österreichischen Expertinnen und Experten die von Berka & Trappel (2019) in einer umfassenden Studie zur Informationsfreiheit in Österreich befragt wurden.

demokratischen Kommunikationsinfrastruktur, für die Vermittlung von Information und damit letztlich auch für die Wahrnehmung der Welt (Just & Latzer 2015). Eingriffe durch Kuratierung und Moderation wirken regulierend auf gesellschaftliche Kommunikation. Öffentlichkeit und Kultur werden durch Inhaltsregulierung auf Internet-Plattformen mitgeprägt. Kuratierung und Moderation sind einflussreich und folgenhaft. Sie können Meinungsfreiheit beschränken und in Konflikt mit ihr geraten.

Vielfach wird deshalb betont, dass dieser starke Einfluss der Plattformen nach einem umsichtigen, ausgewogenen Umgang mit verschiedenen Grundrechten verlangt, insbesondere nach einer adäquaten Berücksichtigung der Meinungsfreiheit, die durch Inhaltsregulierung beschränkt werden kann.[5] Aus dem gesellschaftlichen Ziel einer umsichtigen, ausgewogenen Inhaltsregulierung kann der Bedarf nach einer Kontrolle der Inhaltsregulierung abgeleitet werden. Für Plattformen besteht die Herausforderung in einer Balance auf dem schmalen Grat zwischen Inhaltsregulierung und Meinungsfreiheit, der die Notwendigkeit von Interessenabwägungen einschließt und nach adäquaten Verantwortungsstrukturen verlangt.

Damit stellt sich die Frage, wie die Inhaltsregulierung und ihre Kontrolle organisiert werden können. Welche Grundsätze für die Inhaltsregulierung auf Internet-Plattformen sowie ihre Kontrolle wesentlich sind und welche institutionellen Arrangements sich für Inhaltsregulierung und Kontrolle anbieten, wird in den folgenden Abschnitten analysiert.

2. Governance-Choice: Kontextbedingungen für Inhaltsregulierung

2.1 Analyserahmen

Im zweiten Arbeitsschritt erfolgt ein Assessment der Regulierungsoptionen am Kontinuum zwischen Markt und Staat. Mit Hilfe des Governance-Choice-Ansatzes[6] (vgl. Latzer et al. 2002, 2007; Saurwein 2011; Latzer et al. 2012; Latzer et al. 2019) wird die Herausforderung der Regulierung von Inhalten auf Internet-Plattformen dahingehend überprüft, ob sie sich für Marktlösungen wie die individuelle Selbstorganisation durch Unternehmen oder eine kollektive industrielle Selbstregulierung auf Branchenebene eignet, oder ob ein Bedarf an staatlicher Involvierung bzw. öffentlicher Kontrolle besteht. In der Analyse wird davon ausgegangen, dass die Möglichkeiten und Grenzen für Selbstorganisation und Selbstregulierung u.a. von spezifi-

[5] Siehe dazu vor allem die Positionen von Menschenrechtsorganisationen im Bereich Meinungsfreiheit wie Article 19 und Reporter ohne Grenzen.

[6] Die Methode wurde in verschiedenen Projekten entwickelt und angewendet, beispielsweise zur Analyse von regulatorischen Herausforderungen in eCommerce und Telekommunikation (Latzer et al. 2002, 2003), Inhaltsbewertungen in der audiovisuellen Industrie (Latzer et al. 2007; Saurwein & Latzer 2010), zur Governance von Spam (Just, Latzer & Saurwein 2007), zur Analyse von Jugendschutzfragen in den Bereichen Film, Computerspiele, Telekommunikation und Internet (Latzer et al. 2015) sowie zur Analyse der Governance von Algorithmen im Internet (Latzer, Hollnbuchner, Just & Saurwein 2016; Saurwein, Just & Latzer 2017).

schen Kontextbedingungen abhängig sind, die als mehr oder weniger „ermöglichende Faktoren“ für Marktlösungen wirken können. Die Analyse von Kontextbedingungen soll darüber Aufschluss geben, ob sich Selbstorganisation und Selbstregulierung unter den jeweiligen Bedingungen nachhaltig etablieren und adäquat funktionieren kann, welche Schwächen aus theoretischer Perspektive zu erwarten sind, sowie wo und warum sich gegebenenfalls eine staatliche Involvierung anbietet.

Zu den Kontextbedingungen zählen Charakteristika der Produkte und Industrien ebenso wie regulatorische Kontextbedingungen. Für das Assessment wird ein fixes Set aus folgenden Kontextfaktoren systematisch überprüft:

- der direkte und indirekte Nutzen der Selbstregulierung für die Industrie;
- die Reputationssensitivität der Industrie;
- die Interventionskapazität des Staates;
- die Risiken in Folge von Regulierungsversagen;
- die Intensität der erforderlichen regulatorischen Intervention;
- die Konflikte zwischen privaten und öffentlichen Interessen;
- die Anzahl der Marktteilnehmer, der Grad der Marktfragmentierung und die Wettbewerbsintensität;
- die Verfügbarkeit von Branchenorganisationen für Regulierungsaufgaben;
- die Existenz einer Industriekultur, die auch öffentliche Ziele unterstützt;
- das Niveau und die Formen der staatlichen Involvierung.

Die theoretische Analyse dieser Kontextfaktoren zeigt, welches Potenzial und welche Grenzen die industrielle Selbstregulierung hat und welche Kontextbedingungen für eine öffentliche Kontrolle und eine Involvierung von staatlichen Institutionen sprechen. Im Ergebnis liefert die Analyse von Kontextfaktoren für alternative Regulierungsinitiativen erste Anhaltspunkte für eine mögliche Aufgabenverteilung zwischen staatlichen und privaten Akteuren im Regulierungsarrangement aus theoretischer Perspektive.

2.2 Überprüfung der Kontextfaktoren

Direkter und indirekter Nutzen der Selbstregulierung für die Industrie

Eine Etablierung und der dauerhafte Bestand einer Selbstregulierung sind dann wahrscheinlich, wenn damit ein Nutzen für die Industrie verbunden ist. Die Etablierung ist unwahrscheinlich, wenn der Nutzen der Selbstregulierung geringer ist als der notwendige Aufwand. Zu berücksichtigen ist der direkte Nutzen wie z.B. Erlöse oder der Zugang zu neuen Märkten sowie der indirekte Nutzen, z.B. in Form von Reputationsgewinnen oder indem stärkere staatliche Regulierung durch eine Selbstregulierung verhindert werden kann.

Im Fall der Inhaltsregulierung im Internet steht hohen Kosten vor allem ein indirekter Nutzen gegenüber. Generell ist Inhaltsregulierung im Internet eine sehr schwierige und komplexe Aufgabe, die einen hohen Aufwand und hohe Kosten verursacht – sowohl für die Entwicklung

technischer Systeme zur automatisierten Inhaltskontrolle als auch für den laufenden Betrieb einer manuellen Inhaltskontrolle durch Moderatorinnen und Moderatoren. Technische Systeme verursachen hohe Entwicklungskosten, sind jedoch im Betrieb günstiger und werden aufgrund der Automatisierung und Skalierbarkeit unter ökonomischen Gesichtspunkten gegenüber menschlicher Moderation bevorzugt. Praktisch werden automatisierte und manuelle Systeme meist in Kombination eingesetzt, was die Genauigkeit der Moderation fördert, aber die Kosten zusätzlich erhöht. Alleine Facebook beschäftigt mehr als 20.000 menschliche Moderatorinnen und Moderatoren, die unter hohem Druck täglich eine Vielzahl von Beiträgen begutachten. Skalierbare, automatisierte, technische Lösungen sind im Betrieb kostengünstig, stellen jedoch die größte Gefahr für Beschränkungen der Meinungsfreiheit dar. Umsichtige Einzelfallentscheidungen, an denen mehrere menschliche Moderatorinnen und Moderatoren beteiligt sind, bieten ein hohes Schutzniveau für Meinungsfreiheit, verursachen aber einen besonders hohen Aufwand und Kosten.

Solche Kosten schmälern die Gewinne der meist börsennotierten Plattformunternehmen, deren Vorstände primär den Aktionären verpflichtet sind („Our board of directors is accountable to our shareholders", Zuckerberg 2018). Notwendige Investitionen und Ausgaben stehen unter Rechtfertigungsdruck und müssen durch einen entsprechenden Nutzen begründet sein. Dabei ist der unmittelbare, direkte und monetäre Nutzen aus der Inhaltskontrolle für die Plattformen gering. Inhaltskontrolle generiert in der Regel keine direkten Erlöse. Nur in autoritären Staaten wie z.B. China kann eine Selbstverpflichtung zur Inhaltsregulierung den Zugang zu neuen Märkten ermöglichen und dadurch zu einem direkten monetären Nutzen führen.

Zudem kann durch Inhaltsregulierung der Profit sogar geschmälert werden. Vielfach wird darauf hingewiesen, dass insbesondere jene Inhalte, die hart an der Grenze zu unerlaubten Inhalten liegen (borderline content) eine besonders hohe Nutzeraufmerksamkeit erzielen (vgl. Zuckerberg 2018).[7] Mit der Intensität von Nutzeraufmerksamkeit und Interaktionen steigt auch die Profitabilität solcher Inhalte („metrics go up, money comes in")[8]. Aus ökonomischer Perspektive sinken damit die Anreize für die Plattformen grenzwertige Inhalte zu moderieren und zu löschen. Andererseits kann auch eine hohe Zahl an gesetzeswidrigen und problematischen Inhalten das Nutzererlebnis schmälern, die Plattformen für Teile der Mainstreamnutzer unattraktiv machen und sich negativ auf wichtige ökonomisch relevante Indikatoren wie Nutzerzahl, Verweildauer, Interaktionen und Werbeumsätze auswirken.[9] Inhaltskontrolle kann

[7] Die erhöhte Nutzeraufmerksamkeit hat menschliche und technische Ursachen. Grenzwertige Inhalte, darunter auch Provokation, Hass und Desinformation, lösen Emotionen aus, auf die individuell vergleichsweise stark reagiert wird. Viele Reaktionen durch Klicks und Kommentare führen dazu, dass den jeweiligen Beiträgen im Rahmen der algorithmischen Selektion eine hohe Bedeutung beigemessen wird, die zu einer weiteren Verstärkung der Verbreitung führt (vgl. z.B. Brodnig 2016).

[8] Pointiert formuliert von Zeynep Tufekci, https://twitter.com/zeynep/status/1098607695333130242

[9] „Some amount of inflammatory speech revs people up. But crank the ugliness dial too far, the research shows, and people withdraw. 'There's no incentive for us to have all that stuff (…). People get silenced, they don't engage, and than go off to do something else'." (Van Zuylen-Wood 2019).

für die Plattformbetreiber deshalb einen indirekten Nutzen bieten, wenn es gelingt die Plattformen für möglichst viele Nutzer attraktiv zu gestalten.

Zudem kann ein Nutzen einer Selbstregulierung auch darin bestehen, dass eine stärkere staatliche Regulierung verhindert werden kann.[10] Die Reihe aktueller Bekenntnisse von Plattformbetreibern, allen voran Facebook, zu einem konsequenteren, schärferen Vorgehen gegen illegale Inhalte, Hasspostings und Desinformation dürfte auch dadurch begründet sein, dass Plattformbetreiber einer stärkeren staatlichen Regulierung zuvorkommen bzw. diese verhindern wollen.

In Summe verweist die Analyse auf erhebliche Kosten der Inhaltsmoderation und das Fehlen eines unmittelbaren direkten Nutzens von Inhaltsregulierung durch Plattformbetreiber sowie indirekte Nutzeneffekte, die sich jedoch nicht einfach quantifizieren und den Kosten gegenüberstellen lassen. Unter reinen Kostengesichtspunkten ist Inhaltsregulierung aufgrund hoher Aufwände und potenziellen Beschränkungen der Profitabilität keine Selbstverständlichkeit. Kostenstrukturen sprechen tendenziell für automatisierte, technische Lösungen, mit denen gleichzeitig die Gefahren für Redefreiheit steigen. Der indirekte Nutzen und Anreize für die Industrie zur Selbstregulierung ergeben sich durch die Verbesserung des Nutzungserlebnisses durch adäquate Inhaltsmoderation, durch die Vermeidung von Imageverlusten und wenn durch Selbstregulierung die staatliche Regulierung verhindert werden kann.

Reputationssensitivität der Industrie

Ob freiwillig selbstregulatorische Maßnahmen ergriffen werden, hängt auch von der Reputationssensitivität und drohenden Imageverlusten ab. Je stärker der Erfolg eines Unternehmens von Reputation abhängt, desto stärker sind die Anreize Selbstregulierung einzuführen, um Reputationsverluste zu vermeiden. Eine hohe Reputationssensitivität ist gegeben, wenn Unternehmen eine hohe öffentliche Bekanntheit haben und Themen bearbeiten, denen ein hohes öffentliches Interesse entgegengebracht wird, wenn sie im Endkundenmarkt agieren und Sanktionen durch Konsumenten zu fürchten haben.

Für Internet-Plattformen ist generell von einer hohen Reputationssensitivität auszugehen. Damit bestehen gute Kontextbedingungen für das Zustandekommen von Selbstregulierung. Internet-Inhalte sind in vielfacher Hinsicht ein Thema des öffentlichen Interesses und öffentlicher Besorgnis. Die Verbreitung von illegalen und problematischen Inhalten wie Kindesmissbrauchsdarstellungen, Hassbotschaften und Belästigung wird regelmäßig öffentlich thematisiert und häufig skandalisiert. Gleichzeitig besteht eine hohe öffentliche Aufmerksamkeit hinsichtlich möglicher Beschränkungen der Meinungsfreiheit durch Inhaltsregulierung und Zen-

[10] „In the U.S., the government induces self-regulation largely through the threat of stringent formal rules and costly litigation should industry fail to deliver socially desired outcomes. Industry thus views self-regulation as a preemptive effort to avoid government involvement“ (Newman & Bach 2004, 388). Vgl. auch Wagner (2018a) zum Thema „Ethics as an Escape from Regulation“.

sur. In beiden Themenkreisen sind Organisationen der Zivilgesellschaft etabliert (Bürgerrechtsbewegungen, NGOs), die Fehlverhalten rasch zu einem öffentlichen Thema machen können. Dadurch besteht für Internet-Plattformen die Gefahr von Reputationsverlusten durch negative Berichterstattung. Zudem haben die Plattformen eine enorme Anzahl an Nutzerinnen und Nutzern, sind in der Öffentlichkeit stark präsent und verfügen über eine hohe öffentliche Bekanntheit, die zu einer hohen Anfälligkeit für Reputationsverluste beiträgt.

Des Weiteren agieren Internet-Plattformen im Endkundenmarkt (B-2-C) mit direktem Konsumentenkontakt. Unternehmerisches Handeln, das zu Reputationsverlusten führt, könnte sich in sinkenden Nutzerzahlen und sinkenden Erlösen widerspiegeln. Allerdings ist aufgrund ökonomischer Charakteristika der Plattformökonomie nicht davon auszugehen, dass sich solche Effekte sehr unmittelbar und in geschäftsrelevantem Umfang einstellen. Zum einen wirken die hohe Verbreitung der Plattformen, ihre starke Marktposition, Netzwerkeffekte und die Niedrigkostensituation dem entgegen, dass sich Nutzer bei Fehlverhalten der Plattformen in großer Zahl von den Plattformen abwenden. Dies zeigen beispielsweise die aktuellen Skandale rund um Facebook (z.B. zuletzt Cambridge Analytica), die trotz beachtlicher öffentlicher Aufmerksamkeit und einschlägiger Kampagnen wie „Delete Facebook“ bislang nicht zu einer massenhaften Reduktion von Nutzerzahlen geführt haben. Zudem ist der Geschäftserfolg der Plattformen nicht unmittelbar von Nutzern und Nutzerzahlen abhängig, denn Plattformen agieren auf „zweiseitigen Märkten“. Die Nutzung der Plattformen ist in der Regel gratis und die Finanzierung erfolgt vor allem über Werbung (ggf. auch über Datenhandel). Sinkende Nutzerzahlen könnten sich somit nicht unmittelbar über Verkäufe und Einnahmen auf dem Endkundenmarkt, sondern nur indirekt über sinkende Werbelöse im Geschäftserfolg bemerkbar machen. Verschlechterungen der Reputation am Publikumsmarkt müssen sich nicht unmittelbar direkt im Geschäftserfolg niederschlagen.

In Summe liefert die Analyse Indikatoren, die auf eine hohe Reputationssensitivität der Plattformen hinweisen. Für eine hohe Reputationssensitivität sprechen die hohen Nutzungszahlen, der direkte Kundenkontakt, die hohe Bekanntheit der Plattformen und das starke öffentliche Interesse an Themen der Inhaltsregulierung auf Plattformen. Plattformen können schnell zum Objekt negativer öffentlicher Aufmerksamkeit werden und müssen Reputationsverluste fürchten. Allerdings müssen sich Reputationsverluste aufgrund ökonomischer Charakteristika der Plattformökonomie nicht unmittelbar auf Nutzerzahlen und den Geschäftserfolg der Plattformen auswirken, was die Reputationssensitivität schmälert. Trotzdem bildet Reputation für Internet-Plattformen einen wichtigen Erfolgsfaktor, und es bestehen entsprechende Anreize Selbstregulierung einzuführen, um Reputationsverluste zu vermeiden.

Interventionskapazität des Staates

Die Interventionskapazität des Staates kann sich auf die Etablierung und den Erfolg von Selbstregulierung auswirken. Wenn staatliche Akteure glaubhaft mit der Implementierung staatlicher regulatorischer Lösungen drohen können („credible threats“), kann dies eine

selbstregulatorische Lösung anstoßen, um eine gesetzliche Regelung zu verhindern. Glaubhafte Drohungen können darüber hinaus Anreize schaffen, die Effektivität bestehender Systeme aufrecht zu erhalten oder zu verbessern. Zu untersuchen ist deshalb, über welche Interventionskapazität staatliche Akteure im Regulierungsfeld der Inhaltsmoderation verfügen, um damit einerseits gegen problematische Inhalte vorzugehen und andererseits die Meinungsfreiheit zu schützen. Dies lässt sich u.a. an der Existenz regulatorischer Kompetenzen, der Praktikabilität einer staatlichen Leistungserstellung und dem internationalen Abstimmungsbedarf bei staatlichen Regulierungen festmachen.

Die Schwierigkeiten und Herausforderungen bei der Regulierung des Internet wurden vielfach analysiert und diskutiert. Generell ist die Interventionskapazität staatlicher Akteure im Internetbereich beschränkt. Staatliche Eingriffe in Medien sind aufgrund der grundrechtlich zugesicherten Meinungsfreiheit grundsätzlich höchst sensibel. Entsprechend werden selbstregulatorische Lösungen in westlichen Demokratien meist gegenüber staatlicher Intervention bevorzugt. Praktisch würde eine rein staatliche Kontrolle von Inhalten im Internet einen erheblichen Aufwand verursachen. Die Bewältigbarkeit dieses Kontrollaufwandes spricht dafür, dass Inhalte im Internet primär von Dienstanbietern selbst kontrolliert werden. Die staatliche Interventionskapazität im Bereich der Inhaltsregulierung divergiert aufgrund der Haftungsregelungen nach Akteuren und Arten von Inhalten, die verbreitet werden. Sie ist vor allem bei nutzergenerierten Inhalten niedrig. Beschränkungen für nationale staatliche Interventionsmöglichkeiten bestehen aufgrund des globalen Charakters des Internet. Ein Teil der problematischen Webinhalte stammt aus dem Ausland und entzieht sich der nationalen Regulierung und Strafverfolgung. Zudem sind aufgrund der Globalität der Kommunikationsmärkte in vielen Fällen internationale regulatorische Reaktionen gefragt. Es wird angenommen, dass Selbstregulierung dafür flexibler und anpassungsfähiger ist als staatliche Regulierung (vgl. Latzer et al. 2015).

Solche generellen Beschränkungen der staatlichen Interventionsmöglichkeiten im Internet widerspiegeln sich zum Teil auch bei der Inhaltsregulierung auf Internet-Plattformen:

Auch bei Internet-Plattformen gelten staatliche Eingriffe in Medien aufgrund der grundrechtlich gesicherten Meinungsfreiheit als höchst sensibel. Entsprechend wird auch bei der Inhaltsregulierung auf Internet-Plattformen grundsätzlich Selbstregulierung gegenüber staatlichen Interventionen bevorzugt. Dies zeigt sich z.B. daran, dass staatliche oder supranationale Einrichtungen wie die EU-Kommission in verschiedenen Regulierungsbereichen (Hassreden, Desinformation, Jugendschutz) Selbstregulierung initiieren.

Andererseits bestehen für Staaten grundsätzlich Pflichten zum Schutz von Grundrechten und regulatorische Kompetenzen zur Definition von unrechtmäßigen Inhalten und zur Durchsetzung solcher Verbote. Geltende straf- und zivilrechtliche Beschränkungen umfassen in Österreich beispielsweise das Verbot von Kindesmissbrauchsdarstellungen, harter Pornographie,

nationalsozialistischer Wiederbetätigung, Verhetzung sowie den Schutz von Persönlichkeitsrechten vor Beleidigungen, Mobbing und Verletzungen der Privatsphäre. Für Strafverfolgungsbehörden besteht die Verpflichtung, gegen verbotene Inhalte vorzugehen, sowohl offline als auch online – und somit auch auf Internet-Plattformen.

Schwierigkeiten bereitet jedoch die Rechtsdurchsetzung bei unrechtmäßigen Inhalten. So wurde kritisiert, dass soziale Netzwerke ihren Pflichten zur Löschung von strafbaren Inhalten nur sehr unzureichend nachkommen.[11] Um die Rechtsdurchsetzung zu verbessern wurde in Deutschland 2018 das Netzwerkdurchsetzungsgesetz[12] in Kraft gesetzt. Betreiber sozialer Netzwerke müssen rechtswidrige Inhalte unter Androhung hoher Geldbußen innerhalb definierter Zeitgrenzen löschen. Zu Problemen der Rechtsdurchsetzung kann es auch bei Beleidigungen und anderen Eingriffen in die Persönlichkeitsrechte kommen. Diese sind durch die Extraterritorialität und Grenzenlosigkeit des Internet und potenziell erheblichen Kosten der Rechtverfolgung bedingt (Berka & Trappel 2019, 92). Die Verfolgung illegalen Verhaltens wird auch dadurch erschwert, dass die Benutzung von erfundenen Identitäten im Netz üblich ist. Bestrebungen, die Anonymität über die nationale Gesetzgebung aufzuheben, wären jedoch aufgrund der länderübergreifenden Struktur des Internet weitgehend wirkungslos und würden Grundrechte beschränken (Bundesrat 2009).

Wie generell im Internet spricht auch bei Internet-Plattformen die hohe Anzahl an Inhalten, (Postings, Bilder, etc.) gegen eine „do it yourself"-Lösung seitens des Staates und dafür, dass eine notwendige Inhaltsregulierung direkt durch die Plattformbetreiber selbst vorgenommen wird. Oft wird darauf verwiesen, dass die personellen Kapazitäten der staatlichen Strafverfolgungsbehörden zur Bewältigung einer Vielzahl und Vielfalt an Regelverstößen nicht ausreichen. So bestehen beispielsweise nach Einschätzung des Justizministeriums in Österreich „strukturelle Probleme" bei der Verfolgung von strafbaren Hasskommentaren, weshalb 2017 die Einrichtung von fünf zusätzlichen Planstellen für Sonderstaatsanwälte zur Bekämpfung von Hasskriminalität angekündigt wurde.

Ein weiterer wichtiger Faktor für die Interventionskapazität von Staaten ist der internationale Abstimmungsbedarf bei staatlichen Regulierungen. Je höher der internationale Abstimmungsbedarf, desto niedriger die Interventionskapazität. Lange Zeit galt für die Internet-Regulierung die Prämisse, dass die Globalität des Internet auch nach globalen regulatorischen Reaktionen verlangt. Nachdem jedoch internationale Regelungen häufig nicht zustande kommen oder sich als zu wenig wirksam erweisen, sind seit einigen Jahren verstärkte Vorstöße für nationale und europäische Regulierungen zu beobachten, die speziell auf Internet-Plattformen

[11] Vgl. exemplarisch die Kritik von Jugendschutz.net (2018): „Kinder und Jugendliche sind Mobbing, Hass und Gewalt in Social-Media-Diensten oft schutzlos ausgeliefert. Verstöße werden von den jugendaffinen Plattformen zu selten gelöscht."

[12] Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG) (2017).

abzielen – so zum Beispiel auf EU-Ebene im Datenschutz mit der neuen Datenschutzgrundverordnung und mit einem aktuellen regulatorischen Vorstoß zur Bekämpfung terroristischer Inhalte[13] sowie auf nationaler Ebene mit dem Netzwerkdurchsetzungsgesetz und dem Leistungsschutzrecht in Deutschland oder einem neuen Gesetz gegen Desinformation in Frankreich.[14] Auch wenn diese Interventionen noch neu sind und die Wirksamkeit noch nicht geprüft ist, verweisen die Beispiele doch auf die Interventionskapazitäten der Staaten. Aus theoretischer Perspektive sollte der dadurch aufgebaute Druck (credible threats) dazu beitragen, dass die Bereitschaft der Plattformen zu selbstregulatorischen Lösungen steigt. Mit Blick auf die Handlungspraxis der Unternehmen wird jedoch darauf hingewiesen, dass in vielen Risikofeldern bislang nur harte staatliche Regelungen zu Verhaltensänderungen der Internet-Unternehmen geführt haben (Nemitz 2018).

Während bei verbotenen Inhalten also staatliche Interventionskapazitäten vorhanden sind und zunehmend genutzt werden, sind die Interventionskapazitäten von Staaten zum Schutz der Meinungsfreiheit auf Internet-Plattformen stark beschränkt. Das Grundrecht auf Meinungsfreiheit schützt die Meinungsäußerung vor staatlichen Eingriffen,[15] nicht aber vor Beschränkungen durch private Plattformen. Plattformbetreiber können eigene Regeln setzen und durchsetzen. Die Community-Standards und Regelungsmaßnahmen der Plattformbetreiber entziehen sich weitgehend einer staatlichen Kontrolle hinsichtlich von Beschränkungen der Meinungsfreiheit. Lediglich die sogenannte mittelbare Wirkung der Grundrechte könnte staatliche Interventionen in die Privatautonomie der Plattformbetreiber rechtfertigen.[16] Davon wird bislang aber kaum Gebrauch gemacht, und Interventionen erscheinen - wenn überhaupt - dann nur im Fall von Sperrungen adäquat. Hinsichtlich anderer Formen der Inhaltsregulierung wie Kuratierung bestehen keine staatlichen Interventionskapazitäten.

Zusammenfassend zeigt die Analyse der staatlichen Interventionskapazität zur Inhaltsregulierung und zum Schutz der Meinungsfreiheit damit ein sehr komplexes Bild. Annahmegemäß soll eine hohe Interventionskapazität des Staates zu einer höheren Bereitschaft zur Selbstregulierung führen. Praktisch ist die Interventionskapazität staatlicher Akteure im Internetbereich auf Grund fehlender Legitimität (Zensurverbot) und Praktikabilität (Menge der Inhalte) sowie Schwierigkeiten bei der Rechtsdurchsetzung beschränkt. Andererseits bestehen Möglichkeiten und Verpflichtungen zum Vorgehen gegen illegale Inhalte und zum Schutz der Rechte

[13] Siehe Rat der Europäischen Union (2018), Interinstitutional File 2018/0331(COD)

[14] Siehe Proposition de loi relative à la lutte contre la manipulation de l'information. http://www.assemblee-nationale.fr/15/ta/tap0190.pdf

[15] Die Pflicht des Staates zu einer Sicherung der Meinungsfreiheit ist weitgehend darauf beschränkt, diese nicht durch staatliche Regeln zu beschneiden. Kritisiert wird deshalb beispielsweise, dass Regelungen wie das NetzDG in Deutschland dazu geführt haben, dass Plattformbetreiber zur Sicherheit auf Basis von Community-Standards potenziell problematische Inhalte löschen, um nicht mit dem Gesetz in Konflikt zu geraten. So wird durch das Gesetz die Meinungsfreiheit mittelbar beschränkt (vgl. z.B. Krempl 2018).

[16] Sie z.B. eine Darstellung aktueller Fälle bei Engeler (2018), https://netzpolitik.org/2018/meinungsfreiheit-warum-facebook-zu-recht-nicht-an-grundrechte-gebunden-ist/

Dritter, und etliche aktuelle regulatorische Vorstöße verweisen darauf, dass Interventionskapazität und Interventionsmotivation durchaus vorhanden sind. Ob der steigende regulatorische Druck auf die Internet-Plattformen zu einer höheren Bereitschaft zur Selbstregulierung führt, kann noch nicht beurteilt werden. Eine Intensivierung der privaten Selbstregulierung wird zudem hinsichtlich potenziell beschränkender Implikationen für die Meinungsfreiheit kritisch gesehen. Dabei ist es wichtig zu beachten, dass staatliche Interventionskapazitäten zur Sicherung der Meinungsfreiheit auf privaten Plattformen nur sehr beschränkt vorhanden sind. Die Sicherung von Meinungsfreiheit auf den Plattformen obliegt somit primär dem verantwortungsvollen Umgang der Plattformbetreiber selbst und der kritischen Beobachtung durch die Zivilgesellschaft.

Risiken in Folge von Regulierungsversagen

Die möglichen Folgen eines Regulierungsversagens determinieren den Bedarf an einheitlichen Mindeststandards und die Notwendigkeit einer staatlichen Involvierung. Je höher die Risiken infolge einer defizitären Regulierung sind, desto größer ist der Bedarf an einheitlichen Mindeststandards und an staatlicher Involvierung, um diese Mindeststandards flächendeckend durchzusetzen. Als Indikatoren zur Erfassung dienen die Schadenshöhe, die Folgen in einem Einzelfall, in dem die Regulierung versagt, und der Grad der Irreversibilität von negativen Folgen. Je höher und je irreversibler der Schaden, desto schlechter eignet sich die Herausforderung für reine freiwillige Selbstregulierung.

Im Fall der Inhaltsregulierung auf Internet-Plattformen wären die Folgen eines Regulierungsversagens erheblich. Ein gänzlicher Verzicht auf eine Inhaltsregulierung würde zu einer gesellschaftlich unerwünscht hohen Anzahl an illegalen und problematischen Inhalten (Hass, Desinformation, Terror, Kindesmissbrauchsdarstellungen) auf den Internet-Plattformen führen. Hohe Anteile problematischer Inhalte wirken sich negativ auf die Nutzung sowie auf menschliche Beziehungen, Kultur, Öffentlichkeit und Demokratie aus. Aktuelle Debatten um den Einfluss von Manipulation und Desinformation auf den Ausgang von Wahlgängen sind dafür beispielhaft. Zudem bestehen Risiken bei systematischen Schwächen eines bestehenden Systems Inhalte rechtskonform und konsistent zu moderieren. Risiken entstehen, wenn Inhalte willkürlich gefiltert werden, wenn Kuratierung zu systematischen Verzerrungen führt – oder zu überschießenden Beschränkungen durch „overblocking".

Die Folgen einer fehlenden oder mangelhaften Inhaltskontrolle im Einzelfall können nicht pauschal beurteilt werden. Einerseits richtet nicht jeder einzelne illegale oder unerwünschte Inhalt im Internet unweigerlich einen großen Schaden an und es wird zu Recht darauf verwiesen, dass es für liberale Demokratien und Gesellschaften besonders wichtig ist, auch Inhalte in der Grauzone zu tolerieren. Andererseits bestehen immer wieder hohe Risiken, auch in Einzelfällen. So haben Gerüchte und Falschmeldungen zu anti-muslimischer Gewalt in Sri Lanka und zum Genozid der Rohingya in Myanmar beigetragen.[17] Hohe Risiken und Schäden im

[17] Sie z.B. McKirdy (2018), https://edition.cnn.com/2018/11/06/tech/facebook-myanmar-report/index.html

Einzelfall können auch durch Shitstorms, Mobbing oder falsche Anschuldigen entstehen, die für Einzelne mitunter dramatisch verlaufen. Daraus lässt sich ein Bedarf einer Regulierung zum Schutz vor derartigen Effekten im öffentlichen Interesse und zum Schutz von Persönlichkeitsrechten ableiten, die durch verschiedene rechtliche Schranken für die Kommunikation im Internet festgelegt sind.

Intensität der erforderlichen regulatorischen Intervention

Die Notwendigkeit einer staatlichen Involvierung hängt von der Intensität der regulatorischen Intervention ab. Je niedriger die Eingriffsintensität, desto besser die Eignung für eine selbstregulatorische Lösung. Die Intensität wird durch Charakteristika des Regulierungsproblems und die zur Verfügung stehenden Instrumente bestimmt. Indikatoren sind der Umfang eines Eingriffs in die Grundrechte und ökonomische Auswirkungen der regulatorischen Maßnahmen auf Marktakteure. So sind beispielsweise Eingriffe, die lediglich die Transparenz der Produkteigenschaften verbessern, wenig eingriffsintensiv. Eingriffe in den Marktzugang oder in die Preisbildung sind intensiver und haben einen stärkeren Legitimationsbedarf. Je höher der Legitimationsbedarf, desto stärker die Notwendigkeit einer staatlichen Involvierung, um diese Legitimation sicherzustellen.

Im Fall der Inhaltsregulierung im Internet liegt grundsätzlich eine hohe Eingriffsintensität vor, weil damit Eingriffe in die Redefreiheit verbunden sind. Für die Praxis der Inhaltsregulierung im Internet stehen jedoch unterschiedliche Ansätze und Instrumente zur Verfügung, die sich in ihrer Eingriffsintensität voneinander unterscheiden, weil sie das Marktverhalten der Marktakteure in unterschiedlicher Art und Intensität beschränken. Besonders eingriffsintensiv sind technische Filter, mit denen die gesamte Information und Kommunikation auf den Plattformen pauschal kontrolliert wird, und denen automatisierte Sperren von Inhalten und Profilen folgen können. Mit einer Sperre von Profilen kann eine Beschränkung des Marktzugangs verbunden sein. Als etwas weniger eingriffsintensiv gelten Löschungen, die von menschlichen Moderatorinnen und Moderatoren nach Beurteilungen im Einzelfall vorgenommen werden. Eine geringere Eingriffsintensität ist gegeben, wenn Inhalte nicht gelöscht werden, sondern sichtbar bleiben, aber im Rahmen der Kuratierung weniger prominent dargestellt werden. Weiters gibt es auch Formen der Inhaltsregulierung, die auf die Kennzeichnung von potentiell problematischen Quellen und Inhalten und die gezielte Verlinkung von Gegenmeinungen oder ergänzenden Informationen setzen. Solche transparenz-orientierte Ansätze sind weniger eingriffsintensiv als das Sperren oder Löschen von Inhalten.

In der Praxis zeigt sich, dass die Plattformbetreiber verschiedene Lösungsansätze anwenden, die nicht nur „weiche" transparenz-orientierte Ansätze umfassen, sondern auch harte Eingriffe mittels Löschen und Sperren, sowohl durch menschliche Moderation als auch durch technische Lösungen. *Diese Eingriffe begründen die Sichtweise von Plattformen als „Regulatoren" von gesellschaftlicher Kommunikation* (vgl. Bunting 2018 und Klonick 2017, die Plattformen als

„New Governors“ bezeichnet)[18] . Mit dem Umfang dieser Regulierungsrolle und der steigenden Intensität der Eingriffe im Rahmen der Inhaltsregulierung steigen auch der Legitimationsbedarf und die Notwendigkeit einer externen Kontrolle. Besonders eingriffsintensive Maßnahmen sollten aufgrund der fehlenden demokratischen Legitimation nicht von privaten Akteuren ohne adäquate Kontrollmechanismen ergriffen werden.

„We need a new definition that focuses on the particular role intermediaries play in online information markets. This can be thought of as a form of market governance: the design, implementation and enforcement of rules concerning the free flow of information and content exchange. Unlike 'mere conduits', this governance function means content intermediaries cannot be neutral between different types of content. The content and effects of their community standards, personalisation algorithms and commercial policies will inevitably be controversial.“ (Bunting 2018, 11)

Dabei gilt, dass Inhaltsregulierung auf Plattformen in Folge richterlicher oder behördlicher Anordnung den rechtstaatlichen Verfahren folgen und entsprechende Kontrollmechanismen im Instanzenzug vorsehen. Kritisiert werden jedoch die bislang weitgehend fehlenden Kontrollmechanismen zur Inhaltsregulierung durch Plattformbetreiber, die jenseits rechtstaatlicher Anordnungen selbstständig vorgenommen werden. Problematisiert werden u.a. die Beurteilung von urheberrechtsverletzenden Inhalten durch private Betreiber, temporäre Sperren von Accounts, die als Beschränkung des Marktzugangs wirken können, das Löschen von Inhalten durch Algorithmen abseits rechtsstaatlicher Grundlagen, Formen von „overblocking“ und die Bestimmung von „schädlichen, unerwünschten oder anstößigen“ Online-Inhalten – besonders relevant bei den Phänomenen von Hass-Kommentaren und „Fake News“. Dabei liegen z.T. starke regulatorische Interventionen vor, wobei z.B. bei Netzsperren durch private Akteure und ohne richterliche Kontrolle dem Gebot eines rechtsstaatlichen Verfahrens nicht ausreichend Rechnung getragen wird (vgl. Berka & Trappel 2019, 91).

Zusammenfassend zeigt die Analyse eine hohe Intensität der regulatorischen Interventionen im Rahmen der Inhaltsregulierung. Die Kritikpunkte verweisen insbesondere auf die Notwendigkeit adäquater institutioneller Vorkehrungen zur (externen) Kontrolle der Inhaltsregulierung durch Plattformen, um Willkür zu verhindern, die Rechte der Betroffenen zu wahren, faire Verfahren zu ermöglichen und sicher zu stellen, dass Inhaltsregulierung transparent und rechtmäßig erfolgt. Aus der Rolle der Plattformen als Regulatoren öffentlicher, gesellschaftlicher Kommunikation lässt sich ein Bedarf an Verantwortlichkeit und Kontrolle begründen.

Konflikte zwischen privaten und öffentlichen Interessen

Der Grad der Interessenskonflikte zwischen öffentlichen und privaten Industrieinteressen ist mitentscheidend dafür, ob eine selbstregulatorische Lösung angemessen ist. Je niedriger die

[18] We must abandon traditional doctrinal and regulatory analogies and understand these private content platforms as systems of governance (Klonick 2017).

Interessensgegensätze, desto geeigneter sind selbstregulatorische Lösungen. Bei starken Divergenzen zwischen öffentlichen und privaten Interessen ist reine Selbstregulierung durch die Industrie oft nicht angemessen, weil eine einseitige private Interessensdurchsetzung auf Kosten des öffentlichen Interesses zu befürchten ist.

Inhaltsregulierung ist ein heikles Thema, das verschiedene Dimensionen von inhärenten Interessengegensätzen enthält. Dazu zählen Interessengegensätze zwischen unterschiedlichen Plattformnutzern, zwischen jener der Öffentlichkeit und Plattformbetreibern und Wertekonflikte durch die Kollision von Grundrechten.

Interessengegensätze bestehen auf der Ebene von einzelnen Nutzern und Inhalten. Dem individuellen Bedürfnis von Sprechern nach Veröffentlichung spezifischer Beiträge steht das individuelle Bedürfnis anderer gegenüber, die Veröffentlichung bestimmter Inhalte zu beschränken. Vielfach handelt es sich bei umstrittenen Beiträgen um Inhalte in Grauzonen:

> *„A great deal of problematic content sits in the ambiguous territory between disagreeable political speech and abuse, between fabricated propaganda and legitimate opinion, between things that are legal in some jurisdictions and not others. Or they're things some users want to read and others don't." (Electronic Frontier Foundation 2018)*[19]

Daraus entsteht der Bedarf an Institutionen, die mit Grauzonen und Interessengegensätzen adäquat umgehen. „If many cases are in grey zones, our institutions need to be designed for them" (ebd.).

In der überwiegenden Anzahl der Fälle entscheiden über die Angemessenheit der Inhalte auf ihren Plattformen die Plattformbetreiber auf Basis ihrer Community-Standards, mitunter in Reaktion auf Anordnung von Behörden. Plattformen fungieren mit der Beurteilung von Inhalten als Schiedsrichter, die über die Durchsetzung von Nutzerinteressen entscheiden. Plattformen können in diesen Entscheidungen jedoch nicht „neutral" sein. Viele konkrete inhaltliche Entscheidungen bleiben umstritten. Die Inhalte, die Durchsetzung und die Konsequenzen ihrer Regeln und Standards, ihrer Algorithmen, ihrer kommerziellen Praxis wird immer kontrovers sein (Bunting 2018, 11). Die Interessengegensätze unterstreichen den Bedarf an institutionellen Vorkehrungen, um die Rechte der Betroffenen zu wahren, Einspruchsmöglichkeiten vorzusehen und faire Verfahren zu ermöglichen.

Plattformen haben im Rahmen der Inhaltsmoderation ein wirtschaftliches und organisatorisches Interesse an möglichst einfachen, möglichst globalen Standards für die Zulässigkeit von Inhalten, und an möglichst einfachen, skalierbaren, automatisierbaren Verfahren für Inhaltsregulierung. Dem stehen öffentliche Interessen bezüglich der Inhaltsregulierung gegenüber,

[19] McSherry, York & Cohn (2018), https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some

die sich aus nationalen und regionalen Spezifika ergeben, die zu Unterschieden hinsichtlich der Legalität und gesellschaftlichen Akzeptanz von Inhalten führen. Beispielhaft dafür sind Verbote für nationalsozialistische Inhalte in Deutschland und Österreich. Plattformen geraten dadurch in ein Dilemma (vgl. Spielkamp 2018): „(M)aking the law of one country (i.e. the US) the basis of what's admissible would mean that Facebook would be unable to operate in many other countries in the world (don't just think Saudi Arabia, also think Germany or France), but making the laws of all the countries it operates in the basis would either result in no content to look at or no cross-border network. So both options are no-go areas for the company but also champions of free speech." Die Unterschiede zwischen öffentlichen Interessen und privaten Interessen sprechen gegen eine reine Selbstregulierung und für institutionelle Vorkehrungen, die *kontextsensible Ansätze für Inhaltsregulierung* fördern ohne die Funktionsfähigkeit der Plattformen zum Erliegen zu bringen.

Im Rahmen der Inhaltsregulierung kommt es auch zu Normen- und Wertekonflikten, die sich in der Kollision von verschiedenen Grundrechten widerspiegeln. Dabei geht es in der Regel um Kollisionen zwischen dem Recht auf Meinungsfreiheit und anderen Rechten, wie Persönlichkeits- und Urheberschutz. Durch Eingriffe in Form von Sperren oder Löschen kann die Meinungsfreiheit und damit eine Grundrechtsposition beschränkt werden. Werden bestimmte Inhalte hingegen nicht entfernt, kann es zur Beeinträchtigung von Persönlichkeits- und Urheberrechten kommen. In rechtstaatlichen Verfahren erfolgt eine an den Grundrechten orientierte Interessenabwägung, etwa zwischen der Meinungsfreiheit und entgegenstehenden Rechten, wie z.B. dem Recht auf Achtung der Privatsphäre.[20] Aus diesem Interessenkonflikt lassen sich Forderungen ableiten, auch bei den „privaten" Entscheidungen der Internet-Plattform eine umsichtige Interessenabwägung vorzunehmen.

Zusammenfassend verweist die Analyse auf Interessenunterschiede zwischen unterschiedlichen Plattformnutzern und zwischen öffentlichen Interessen und den Interessen der Plattformbetreiber sowie auf Norm-/Wertekonflikte zwischen der Meinungsfreiheit und anderen Grundrechten. Dies verdeutlicht den Bedarf an einem institutionellen Arrangement zur Inhaltsregulierung, das geeignet ist diese Interessengegensätze zu bewältigen. Ein solches Arrangement erfordert institutionelle Vorkehrungen, die geeignet sind mit den Grauzonen zwischen akzeptablen und inakzeptablen Inhalten umzugehen, die Rechte der Betroffenen zu wahren, Einspruchsmöglichkeiten vorzusehen und faire Verfahren zu ermöglichen, die kontextsensible Ansätze für Inhaltsregulierung ermöglichen ohne die Funktionsfähigkeit der Plattformen zu gefährden, und auch bei den „privaten" Entscheidungen der Internet-Plattform eine umsichtige Interessenabwägung mit Blick auf Grundrechte vorzunehmen.

[20] In rechtsstaatlichen Verfahren ist es ständige Judikatur, „dass die einschlägigen Tatbestände im Lichte der Grundfreiheiten und Menschenrechte auszulegen und anzuwenden sind, was in der Regel – vor allem beim zivilrechtlichen Persönlichkeitsschutz – auf eine an den betroffenen Grundrechten orientierte Interessenabwägung zwischen der Meinungsfreiheit und entgegenstehenden Rechten, wie etwa dem Recht auf Achtung der Privatsphäre, hinausläuft" (Berka & Trappel 2019, 21f).

Anzahl der Marktteilnehmer, Grad der Marktfragmentierung & Wettbewerbsintensität

Die Möglichkeiten und die Angemessenheit einer Selbstregulierung werden durch die Anzahl der Marktteilnehmer, den Grad der Marktfragmentierung und die Intensität des Wettbewerbs mitbestimmt. Aus theoretischer Perspektive gilt: Je stärker die Wettbewerbsintensität, desto stärker die Fokussierung auf ökonomischen Erfolg und desto geringer die generellen Anreize, dass neben ökonomischen Zielsetzungen auch soziale Ziele verfolgt werden. Je höher die Anzahl der Markteilnehmer und je fragmentierter die Industrie ist, desto schwieriger ist es, Selbstregulierungsinitiativen mit starker Abdeckung einzuführen und ihren Bestand zu sichern. Je niedriger die Anzahl der Marktteilnehmer und je niedriger die Anzahl der Teilnehmer an einer Selbstregulierung, desto höher die Gefahr von Absprachen und Kartellbildungen und steigender Machtkonzentration.

Die Markt- und Wettbewerbsstrukturen am Markt der Internet-Plattformen sprechen für das Zustandekommen einer Selbstregulierung, begründen jedoch gleichzeitig Gefahren einer zu starken Machtkonzentration.

Am Markt der Internet-Plattformen und Informations-Intermediäre dominieren derzeit einige wenige Anbieter wie Facebook, YouTube und Twitter. Damit bestehen generell gute Marktstrukturvoraussetzungen für die Etablierung einer Selbstregulierung, die sich organisatorisch am leichtesten in Oligopolen einführen lässt. Denn unter wenigen Teilnehmern gelingt es leichter einen Konsens über eine Selbstregulierung zustande zu bringen, speziell dann, wenn sich die Teilnehmer hinsichtlich ihrer Markposition auf Augenhöhe begegnen.

Gleichzeitig bestehen erhebliche Bedenken hinsichtlich der Implikationen einer Selbstregulierung für die Einflussmacht der Plattformunternehmen. Schon heute verfügen die Plattformen über eine starke ökonomische Macht in Kombination mit einem starken Einfluss auf die gesellschaftliche Kommunikation, die immer mehr über ihre Kanäle verläuft. In vielen westlichen Ländern mit starker Internetpenetration sind soziale Netzwerke beispielsweise dabei, traditionelle Medien wie Rundfunk und Print als meistgenutzte Informationsquelle abzulösen. Vielfach wird von einer umfassenden „platform power“ und einer „digital dominance“ (Moor & Tambini 2018) gesprochen, die neben ökonomischem Erfolg und dem Einfluss auf gesellschaftliche Kommunikation auch eine immer stärkere politische Macht der Plattformen umfasst (Helberger 2018). Mark Zuckerberg selbst meint: “Facebook is now more like a government than a traditional company”.[21]

Eine verstärkte Übertragung von Verantwortung für Inhaltsregulierung auf Plattformbetreiber würde zu einem weiteren Zuwachs von Macht und Einfluss der Plattformbetreiber führen. Insbesondere eine verstärkte Haftung der Plattformbetreiber für unrechtmäßige Inhalte kann aus dieser Perspektive problematisiert werden, weil diese zu einem noch stärkeren Einfluss auf Inhalte und zu einer Verstärkung der politischen Machtposition der Plattformen beiträgt.

[21] Zitiert nach Ezra Klein, https://www.youtube.com/watch?v=xQobbtRnozk

Hinzu kommen potenziell problematische Effekte für den Plattformwettbewerb. Strengere Haftungsregeln und höhere Anforderungen an Inhaltsmoderation können eher von etablierten, großen und ressourcenstarken Unternehmen als von neuen, kleinen Start-ups umgesetzt werden (Keller 2018). Hohe Anforderungen können den Plattformwettbewerb verringern und große bestehende Plattformen weiter stärken.

Zusammenfassend bieten Marktstrukturen gute Voraussetzungen für das Zustandekommen einer Selbstregulierung, begründen jedoch gleichzeitig Gefahren einer zu starken Machtkonzentration. Eine umfassende und alleinige Umsetzung und Kontrolle der Inhaltsregulierung durch Plattformbetreiber stellt deshalb keine Ideallösung dar. Daraus lässt sich z.B. ableiten, dass ein Bedarf an einer stärkeren Verteilung von Verantwortlichkeit (distributed accountability) in der Inhaltsregulierung besteht – zum Beispiel indem durch „checks and balances" die operative Inhaltsregulierung, Berufungsmöglichkeiten und die Aufsicht strukturell getrennt werden.

Verfügbarkeit von Branchenorganisationen zur Wahrnehmung regulatorischer Aufgaben

Die Umsetzung von Selbstregulierung erfolgt vielfach durch Organisationen wie Branchenverbände, die idealerweise über ausreichend Reichweite, Kapazitäten und Autorität verfügen, um regulatorische Aufgaben wahrzunehmen. Die Verfügbarkeit von entsprechend anerkannten Organisationen ist ein wichtiger Faktor für die Einführung und den nachhaltigen Bestand von kollektiver Selbstregulierung. Zu überprüfen ist deshalb, ob die Branchen über geeignete Organisationen verfügen.

Für die Internetwirtschaft haben sich auf internationaler und nationaler Ebene zahlreiche Verbände gebildet, um die Industrieinteressen zu vertreten. Zu diesen zählen beispielweise die Global Network Initiative (GNI), die Computer and Communications Industry Association (CCIA), Digital Europe, die European Digital Media Association (EDiMA), die Interactive Digital Advertising Alliance (EDAA) oder das Interactive Advertising Bureau (IAB) Europe. Auch die großen Internet-Plattformen sind Mitglied in Verbänden, die mit ihren Geschäftsfeldern in Verbindung stehen. Einige dieser Verbände haben auch Selbstregulierungsinitiativen gestartet, z.B. im Werbesektor.[22] An anderen Initiativen, z.B. zum Jugendschutz[23], beteiligen sich die Unternehmen direkt und nicht über einen Branchenverband. Dies gilt auch für neuere Initiativen gegen Hassrede und Desinformation, die zuletzt auf Betreiben der EU-Kommission eingerichtet wurden. Im Rahmen der Global Network Initiative (GNI) wurden Grundsätze (GNI Principles)[24] beschlossen, die u.a. einen Verhaltenskodex umfassen, der festlegt, wie Netzwerke mit staatlichen Löschansuchen umgehen sollen.

---

[22] Zum Beispiel ein Selbstregulierungssystem für Online Behavioural Advertising im Rahmen der European Interactive Digital Advertising Alliance (EDAA) und Spezifikationen für einen Do-Not-Track-Standard im Rahmen des Word Wide Web Consortiums (W3C).

[23] Siehe zum Beispiel Verhaltenskodizes der Freiwilligen Selbstkontrolle Multimedia (FSM) in Deutschland.

[24] Siehe https://globalnetworkinitiative.org/gni-principles/

Die Beispiele zeigen, dass im Sektor der Internet-Plattformen Strukturen für Selbstregulierung und Erfahrungen mit Selbstregulierungsinitiativen bestehen. Aufgrund der Marktstrukturen mit wenigen großen Anbietern können auch Initiativen mit direkter Beteiligung eine starke Reichweite im Markt erzielen. Branchenorganisationen sind deshalb keine conditio sine qua non für erfolgreiche Selbstregulierung im Plattformbereich. Gleichzeitig gibt es erhebliche Kritik und berechtigte Zweifel an der Funktionsfähigkeit der Selbstregulierung. Bei etlichen der Initiativen handelt es sich um Kodizes und Bekenntnisse ohne Durchsetzungsmechanismus und Erfolgskontrolle. Dies führt zur Kritik, dass Selbstregulierung wie auch Ethik-Initiativen lediglich dazu dienen, eine strengere Regulierung zu verhindern (vgl. z.B. Wagner 2018a). Zudem waren speziell die europaweiten Industrieinitiativen im Bereich des Datenschutzes nicht nachhaltig erfolgreich und konnten keinen ausreichenden Schutz bieten. Bei inhaltlich orientierten Initiativen gegen Hassrede und Desinformation wird kritisiert, dass Plattformen nun zu Richtern über Inhalte werden und die Meinungsfreiheit einschränken.

Darüber hinaus sind Selbstregulierungsinitiativen wiederholt gescheitert. So waren zentrale Plattformunternehmen 2013 nicht bereit, einen geplanten Kodex für soziale Netzwerke zu unterzeichnen und heikle Themen wie Daten- und Jugendschutz mit einer Selbstverpflichtung zu regeln. Diesbezügliche Stellungnahmen der Plattformbetreiber verdeutlichen exemplarisch, dass die Ambitionen zur Teilnahme an nationaler Selbstregulierung gering sind. „Es liegt in der Natur unserer international ausgerichteten Dienste, dass wir uns nicht in jedem Land an nationalen Programmen zur Selbstregulierung beteiligen können", sagte ein Google-Sprecher. Ein Facebook-Sprecher sagte: „Wir haben weltweit eine Reihe von Initiativen zur Selbstregulierung und müssen diese miteinander in Einklang bringen, bevor wir regionale Kodizes unterzeichnen."[25]

Zusammenfassend zeigen die Beispiele, dass im Sektor der Internet-Plattformen Strukturen für Selbstregulierung und Erfahrungen mit Selbstregulierungsinitiativen bestehen. Praktisch sind die bestehenden Initiativen jedoch bislang auf bestimmte Risiken (Werbung, Jugendschutz) beschränkt und nicht durchwegs erfolgreich. Aufgrund globaler Interessen und organisatorischer Barrieren bestehen Widerstände der Industrie und berechtigte Zweifel am Zustandekommen von nationalen Lösungen im Wege einer Selbstregulierung.

Industriekultur, die öffentliche Ziele unterstützt

Die Wahrscheinlichkeit, dass Selbstregulierung zustande kommt, effektiv funktioniert und öffentliche Ziele umsetzt, hängt auch von der Industriekultur ab. Eine existierende Sensitivität in einer Branche für öffentliche Interessen und eine Tradition der Kooperation und des Interessenausgleichs fördert das Zustandekommen von Selbstregulierung.

[25] DW (2013), https://www.dw.com/de/selbstregulierung-von-online-netzwerken-gescheitert/a-16795312

Die Etablierung einer Selbstregulierung ist deshalb wahrscheinlicher in reifen Industriesektoren mit gleichgesinnten Marktteilnehmern, die sich auf Augenhöhe begegnen. Märkte, in denen Internet-Plattformen zum Einsatz kommen, sind hingegen entweder neu und experimentell oder die Plattformen fordern als Newcomer etablierte Marktstrukturen, Anbieter und Geschäftsmodelle heraus. Dabei suchen sie bewusst nach neuen Wegen zum Erfolg und lassen sich ihre Marktchancen nicht durch freiwillige Selbstbeschränkungen schmälern. „Move fast and break things" galt lange als Motto von Facebook und steht exemplarisch für eine Industriekultur („Silicon Valley Culture"), die nur wenig für Beschränkungen im öffentlichen Interesse übrig hat.

Erst im Zuge der aktuellen Skandale rund um den Datenschutz, erhöhtem öffentlichen Druck und im Verlauf der parlamentarischen Anhörungen kommt es verstärkt zu Schuldeingeständnissen und Entschuldigungen der verantwortlichen Plattformbetreiber. Ob sich damit auch ein Wandel in der Industriekultur ankündigt, bleibt abzuwarten.

Adäquates Niveau und passende Form staatlicher Involvierung

Die Effektivität von Regulierungsarrangements hängt auch davon ab, ob ein passender Mix aus privaten und öffentlichen Beiträgen gefunden wird. So ist der Begriff Selbstregulierung in vielen Fällen praktisch eine Fehlbezeichnung, weil die meisten Selbstregulierungsinitiativen nicht völlig ohne staatliche Involvierung in ihrer Entwicklung oder im operativen Betrieb auskommen. Ein erfolgreiches institutionelles Design hängt davon ab, ob ein passendes Niveau und angemessene Formen der staatlichen Involvierung gefunden werden.

Die vielfältige Kritik an der Inhaltsregulierung auf Internet-Plattformen ist ein Indikator dafür, dass ein angemessener Mix bislang noch nicht gefunden wurde. Um einen angemessenen Mix zu bestimmen, ist es hilfreich auf die oben analysierten Kontextfaktoren zu achten und daraus entsprechende Schlussfolgerungen zu ziehen.

2.3 Fazit und Schlussfolgerungen zu Kontextbedingungen für Inhaltsregulierung

Zusammenfassend legt die Analyse von Kontextfaktoren für Inhaltsregulierung auf Internet-Plattformen ein Regulierungsarrangement nahe, das in der operativen Praxis von den Plattformen selbst betrieben wird und gleichzeitig unter einer öffentlichen Kontrolle steht.

Für Selbstorganisation durch Unternehmen und Selbstregulierung auf Branchenebene spricht vor allem, dass staatliche Eingriffe in Medieninhalte grundrechtsproblematisch sind und deshalb Selbstregulierung zu bevorzugen ist. Für Selbstregulierung sprechen zudem die immense Zahl an Inhalten im digitalen Umfeld, die nur von den Plattformbetreibern selbst moderiert werden kann, der indirekte Nutzen einer adäquaten Inhaltsregulierung für die Industrie (Image, Nutzererlebnis), die Reputationssensitivität der stark exponierten Plattformunternehmen und die oligopolistische Marktstruktur, die die Einführung einer Selbstregulierung erleichtert.

Gleichzeitig verdeutlicht die Analyse, dass Inhaltsmoderation nicht ausschließlich der Industrie überlassen werden sollte und eine öffentliche Kontrolle ratsam erscheint. Im Detail lassen sich aus den Analysen der Kontextfaktoren folgende zentrale *Schlussfolgerungen über Bedarf und Ansatzpunkte einer öffentlichen Kontrolle* ableiten:

Internet-Plattformen agieren als Regulatoren privater und öffentlicher Kommunikation, verwenden dabei zum Teil technische Instrumente in Form von Software mitunter auf Basis künstlicher Intelligenz, um Plattformen zu kontrollieren, und problematische Inhalte zu identifizieren und zu löschen. Diese Praxis bedeutet eine hohe regulatorische Eingriffsintensität im grundrechtssensiblen und demokratierelevanten Bereich gesellschaftlicher Kommunikation. Aus dieser Rolle der Plattformen als Regulatoren lassen sich *Forderungen nach Verantwortlichkeit und nach externer Kontrolle* ableiten, um Willkür zu verhindern, die Rechte der Betroffenen zu wahren, faire Verfahren zu ermöglichen und sicherzustellen, dass Inhaltsregulierung transparent und rechtmäßig erfolgt.

Inhaltsregulierung durch Internet-Plattformen ist mit einer Vielzahl an Interessenkonflikten konfrontiert. Diese ergeben sich zwischen unterschiedlichen Plattformnutzern, zwischen öffentlichen Interessen und den privaten Interessen der Plattformbetreiber, und zwischen unterschiedlichen Grundrechten, die von Inhaltsregulierung betroffen sind. Daraus entsteht der *Bedarf an einem institutionellen Arrangement für Inhaltsregulierung, das geeignet ist diese Interessengegensätze zu bewältigen*. Dazu zählen Vorkehrungen, um mit den Grauzonen zwischen akzeptablen und inakzeptablen Inhalten umzugehen und eine faire Interessenabwägung zu ermöglichen, beispielsweise durch Einspruchs-/Berufungsmöglichkeiten und breite Stakeholdereinbindung.

Internet-Plattformen verfügen bereits heute über eine hohe Marktmacht, starken Einfluss auf private und öffentliche Kommunikation und zunehmend politische Macht. Deshalb bestehen erhebliche Bedenken, wenn Plattformen im Rahmen der Inhaltsregulierung noch weitere Kompetenzen, Verpflichtungen und Verantwortung inklusive Haftung akkumulieren, die zu einer weiteren Verstärkung der Machtpositionen der Plattformen beiträgt. Zwar werden Plattformen regelmäßig zu mehr Verantwortung angehalten. Eine umfassende und alleinige Umsetzung und Kontrolle der Inhaltsregulierung durch Plattformbetreiber stellt jedoch aufgrund zunehmender Verantwortungs- und Machtakkumulation keine Ideallösung dar. Es besteht ein *Bedarf an einer stärkeren Verteilung von Verantwortlichkeit und der Implementierung geeigneter „checks and balances“*, beispielsweise indem Inhaltsregulierung, Beschwerdeverfahren und die Aufsicht strukturell voneinander getrennt werden.

Die Interessen der Internet-Plattformen sprechen tendenziell für möglichst automatisierte Inhaltsmoderation, möglichst internationale Standards und globale institutionelle Arrangements zur Inhaltsregulierung. Automatisierte und globale Lösungsansätze sind organisato-

risch leichter umsetzbar und kostengünstiger als differenzierte Ansätze, gehen jedoch auf Kosten von regionalen Interessen, Kontextsensitivität und umsichtigen Abwägungen im Einzelfall. Eine zentrale Herausforderung für institutionelle Arrangements besteht darin, *kontextsensible Ansätze für Inhaltsregulierung zu ermöglichen*, ohne die Funktionsfähigkeit der Plattformen zu gefährden.

Plattform-Unternehmen sind Teil einer Industriekultur, die gesamtgesellschaftlichen Zielen und Interessen bislang wenig Beachtung schenkt und nur wenig für Beschränkungen im öffentlichen Interesse übrig hat. Auch der erhebliche Aufwand für Inhaltsregulierung, die ökonomischen Ziele der Konzerne, Verpflichtungen gegenüber Aktionären und die schwachen Interventionskapazitäten des Staates (credible threats), die auf verbotene Inhalte beschränkt sind, sprechen tendenziell dagegen, dass Selbstregulierung konsequent implementiert wird.

Kosten-Nutzen-Relationen sprechen dagegen, dass von den Plattformen sehr umsichtige, abwägende und damit kostenintensive Ansätze der Inhaltsregulierung verfolgt werden, die einen vorsichtigen Umgang mit Meinungsfreiheit ermöglichen. Staatliche Interventionskapazitäten zur Sicherung der Meinungsfreiheit auf privaten Plattformen sind kaum vorhanden. Die Sicherung von Meinungsfreiheit auf den Plattformen obliegt primär dem verantwortungsvollen Umgang der Plattformbetreiber selbst und der kritischen Beobachtung durch die Zivilgesellschaft, der entsprechende Kontrollmöglichkeiten eingeräumt werden sollten.

Zusammenfassend lassen sich daher folgende Schlussfolgerungen für Vorkehrungen im öffentlichen Interesse ableiten. Es besteht ein Bedarf an …

- Vorkehrungen im Rahmen der Inhaltsregulierung, um Willkür zu verhindern, die Rechte der Betroffenen zu wahren, faire Verfahren zu ermöglichen und sicherzustellen, dass Inhaltsregulierung transparent, rechtmäßig und verhältnismäßig erfolgt, z.B. durch interne Verantwortlichkeit und externe Kontrolle/Aufsicht;
- Vorkehrungen, um Interessengegensätze zu bewältigen und eine faire Interessenabwägung zu ermöglichen, beispielsweise durch Einspruchsmöglichkeiten und breite Stakeholder-Einbindung;
- Vorkehrungen zur stärkeren Verteilung von Verantwortlichkeit sowie „checks and balances“, z.B. durch strukturelle Trennung von operativer Inhaltsregulierung, Beschwerdeverfahren und Aufsicht;
- Vorkehrungen für kontextsensible Ansätze für Inhaltsregulierung, ohne die Funktionsfähigkeit der Plattformen zu gefährden;
- Vorkehrungen für Transparenz und Kontrollmöglichkeiten für Experten oder die Zivilgesellschaft zur Sicherung von Meinungsfreiheit durch Verhinderung von Willkür, Verzerrungen, Overblocking bei der Inhaltsmoderation;

Generell ist bei Governance-Arrangements mit staatlichen und privaten Beiträgen zu beachten, dass Kompetenzen und Verantwortlichkeiten klar zugeordnet sind (Saurwein 2011, 352).

3. Inhaltsregulierung auf Internet-Plattformen: Status quo in Europa und Österreich

Der folgende Abschnitt bietet eine Beschreibung des Status quo der aktuellen Regulierungspraxis (Wie wird aktuell reguliert?). Diese umfasst zuerst die Maßnahmen zur Inhaltsregulierung, die von Internet-Plattformen gesetzt werden, im Rahmen unternehmerischer Selbstorganisation oder durch Selbstregulierung auf Branchenebene. Im Anschluss erfolgt ein Überblick zu den Eckpfeilern des aktuellen geltenden Regulierungsrahmens in Österreich und der Europäischen Union. Ergänzt werden diese Erhebungen durch eine Analyse zu Schwächen im aktuellen Regulierungsarrangement anhand ausgewählter Kritikpunkte, die im Fachdiskurs geäußert werden (Welche Defizite zeigt das aktuelle Regulierungssystem?). Im Ergebnis liefert Abschnitt 3 einen Überblick zum Stand und zur Kritik der aktuellen Regulierungslage als Begründung für den anstehenden Handlungs-/Reformbedarf.

Inhaltsregulierung auf Internet-Plattformen ist eine komplexe Angelegenheit, die operativ durch einen Mix aus technischer und menschlicher Kontrolle umgesetzt wird, und die auf staatlichen/rechtlichen Vorgaben sowie auf den plattformeigenen Regelungen im Rahmen der *Selbstorganisation* basiert. Hinzu kommen einige neuere Initiativen für kollektive Selbstverpflichtungen, denen sich mehrere wichtige Plattformen im Rahmen einer *Selbstregulierung* angeschlossen haben und die vor allem auf Drängen von Regulierungseinrichtungen wie der EU-Kommission zustande gekommen sind.

3.1 Selbstorganisation: Praxis der Inhaltsregulierung durch die Plattformbetreiber

Im Zentrum der Inhaltsregulierung stehen die Plattformen selbst. Die Plattformbetreiber entwickeln Gemeinschaftsrichtlinien (Community-Standards), um „problematische" Inhalte auf ihren Plattformen zu verhindern und positive Nutzererlebnisse zu fördern, und sie implementieren komplexe Systeme zur Kontrolle von Inhalten, um die unerlaubten Beiträge zu identifizieren, zu filtern, zu löschen oder im Rahmen der Kuratierung weniger prominent darzustellen. Diese Filterung und Kuratierung erfolgen mittels automatisierter Verfahren und menschlicher Kontrollen, die auch als Inhaltsmoderation (content moderation) bezeichnet werden. Die wesentlichen Komponenten der Inhaltsregulierung durch Plattformen werden im Folgenden kurz beschrieben.

Gemeinschaftsstandards

Plattformbetreiber streben danach, ihre Dienste attraktiv zu gestalten und vor Missbrauch zu schützen, und legen Standards dafür fest, welche Inhalte auf den Plattformen gestattet sind und welche nicht. Diese Standards gelten zumeist weltweit und für alle Arten von Inhalten. Sie umfassen in der Regel Verbote für illegale Inhalte, gehen aber häufig darüber hinaus und

setzen weitere plattformspezifische Standards, mit denen festgelegt wird, welche Inhalte nicht akzeptabel sind. So reguliert bzw. untersagt beispielsweise Facebook mit den Gemeinschaftsstandards[26] u.a. Nacktheit, Pornographie, Nacktdarstellung von Kindern und deren sexuelle Ausbeutung, sexuelle Kontaktaufnahme, Bullying, Belästigung, Verletzungen der Privatsphäre und Persönlichkeitsrechte in Bezug auf Bilder, Fake-Accounts, Falschmeldungen, Hassrede, Spam, Gewalt, glaubhafte Gewaltabsichten, die Begünstigung und Propagierung von Verbrechen, Suizid und Selbstverletzung, terroristische Propaganda und Verletzungen von geistigen Eigentumsrechten.

Die Gemeinschaftsstandards fungieren für die Plattformen als normativer Referenzrahmen für Maßnahmen im Rahmen der Inhaltsmoderation. Die Regeln sind im Sinne eines „Hausrechts" legitim, denn grundsätzlich liegt es im Ermessen der Plattformen zu bestimmen, welche Arten von Inhalten sie tolerieren, solange die Inhalte nicht gegen geltendes Recht verstoßen. Kritisiert wird jedoch, dass die Standards zum Teil überschießend sind und die freie Meinungsäußerung beschränken. So wird zum Beispiel das Verbot von jeglichen Nacktbildern auf Facebook als überzogen kritisiert, z.B. weil es künstlerische Darstellungen unterbindet (censorship curates culture)[27]. Zudem sind die Regeln so offen formuliert und allgemein gehalten, dass selbst Nutzer ohne böse Absichten Probleme damit hätten, sich an alle Vorschriften zu halten. Des weiteren wird kritisiert, dass die global geltenden Inhaltsstandards im Wesentlichen US-amerikanische Wertvorstellungen widerspiegeln und zu wenig sensibel für andere nationale Kontexte sind.

Meldestellen und Trusted Flaggers

Die Internet-Plattformen betreiben verschiedene Kanäle, um Meldungen über problematische Inhalte zu empfangen. Individuelle Nutzer können einzelne Beiträge mittels Melde-Button oder über Online-Formulare direkt melden, wenn sie vermuten, dass damit Verstöße gegen Gemeinschaftsregeln oder Rechtsverletzungen vorliegen. Problematisiert wird, dass Meldeprozesse zum Teil wenig benutzerfreundlich sind,[28] dass Nutzermeldungen häufig zu unkonkret erfolgen, und dass die Meldesysteme der Plattformen gezielt von organisierten Gruppen genutzt werden, um Personen und Inhalte zu diskreditieren, Sperrungen zu erwirken und um sich mitunter auch Vorteile im ökonomischen Wettbewerb zu verschaffen.

Neben diesem direkten Kanal für Nutzer, kann eine Meldung auch über spezielle öffentliche oder zivilgesellschaftliche Meldestellen erfolgen. Zudem betreiben die Internet-Plattformen sogenannte „Trusted Flagger"-Programme. Personen oder Organisationen mit „Trusted Flag-

[26] https://de-de.facebook.com/communitystandards/introduction

[27] York & Wagenknecht (2016), https://re-publica.com/de/session/nudes-and-n00dz

[28] Diese Kritik wurde z.B. an den Prozessen zur Meldung von Vergehen nach dem Netzwerkdurchsetzungsgesetz in Deutschland erhoben. Dabei soll von Nutzern angegeben werden, welcher konkrete rechtliche Tatbestand vorliegt, was für juristische Laien schwer zu beurteilen ist.

ger"-Status werden von den Plattformen als besonders vertrauenswürdig eingestuft. Ihre Meldungen werden von den Inhaltsmoderatoren deshalb prioritär überprüft. Die Prüfung erfolgt nach denselben Standards wie bei Inhalten, die von anderen Nutzern gemeldet werden.

In Österreich sind mehrere Meldestellen etabliert: Direkt beim Innenministerium sind Meldestellen für Kinderpornographie und nationalsozialistische Wiederbetätigung eingerichtet. Die *Stopline* ist eine Initiative der österreichischen Internet-Service-Provider und bearbeitet ebenfalls Meldungen zu illegalen Inhalten in den Bereichen Kindesmissbrauchsdarstellungen und nationalsozialistische Wiederbetätigung. Die *ZARA Beratungsstelle #GegenHassimNetz* informiert und unterstützt Opfer von Rassismus und Hassbotschaften und verfügt über einen „Trusted Flagger"-Status. Der *Internet Ombudsmann* bietet kostenlose Schlichtung und Beratung für internetbezogene Themen wie Online-Shopping, Urheberrecht, Datenschutzrecht, Recht am eigenen Bild und Persönlichkeitsrechte an.

Dem teils öffentlichen, teils zivilgesellschaftlich organisierten Meldewesen in Österreich wird im Allgemeinen eine effiziente Funktionsweise zu attestiert, wobei die Zusammenarbeit der Meldestellen mit Strafverfolgungsbehörden weiter verbessert werden sollte (Berka & Trappel 2019, 92). Die EU-Kommission konstatiert, dass die verstärkte Zusammenarbeit zwischen Plattformen und Organisationen der Zivilgesellschaft im Rahmen des Verhaltenskodex zur Bekämpfung illegaler Hassreden zu stichhaltigeren Meldungen, einer wirksameren Behandlung und besseren Ergebnissen bei der Reaktion auf Meldungen geführt hat (Europäische Kommission 2017).[29] Doch auch Meldestellen mit Trusted-Flagger Status beklagen, dass für sie nicht immer klar nachvollziehbar ist warum gemeldete Inhalte von den Plattformen entfernt werden oder nicht. Dies verweist auf Transparenz- und Kommunikationsprobleme an den Schnittstellen zwischen Meldestellen und Plattformen.

Trotzdem fungieren die national verankerten Melde- und Beratungsstellen als niedrigschwellige Anlaufstellen für Nutzer und tragen dazu bei, dass potenziell problematische Beiträge mit der notwendigen regionalen Kontextsensibilität begutachtet werden.

Menschliche Moderation

Die Durchsetzung der Gemeinschaftsstandards erfolgt durch die Filterung problematischer Inhalte, die bislang unter anderem durch menschliche Kontrollen oder Nachkontrollen umgesetzt wird. Die menschlichen Moderatorinnen und Moderatoren bearbeiten Beiträge, die von Nutzern, Behörden oder automatisierten Systemen als problematisch angesehen und gemeldet werden.

Im Rahmen der Inhaltsmoderation betreibt alleine Facebook eigenen Angaben zu Folge zehn Büros in sechs Ländern und beschäftigt über 30.000 Menschen (Zuckerberg 2018), davon etwa 15.000 Moderatorinnen und Moderatoren (Ash et al 2019, 10) die *täglich über zwei Millionen*

[29] Europäische Kommission (2017), http://europa.eu/rapid/press-release_IP-17-1471_de.htm

Beiträge begutachten. Facebook hat die Größe seines Moderationsteams damit innerhalb von 2 Jahren verdoppelt. YouTube verfügt über 10.000 Content Reviewer (Caplan 2018, 24).

Sowohl die Plattformbetreiber selbst als auch Kritiker problematisieren die mitunter fehlende Genauigkeit (accuracy) der menschlichen Inhaltsregulierung. Laut Mark Zuckerberg (2018) werden aktuell für mehr als ein Zehntel der gesichteten Inhalte falsche Entscheidungen durch die menschlichen Moderatorinnen und Moderatoren getroffen.[30] Auch Kritiker verweisen darauf, dass soziale Netzwerke wie Facebook oder Twitter nur einen Bruchteil derjenigen Inhalte löschen, die strafbar oder rechtswidrig sind, und gleichzeitig häufig Postings und Inhalte entfernen, die erkennbar rechtmäßig sind. Die Kriterien, nach denen Facebook & Co. ihre Löschentscheidungen treffen, gelten als intransparent und nicht nachvollziehbar (vgl. z.B. Stadler 2018). Nach jahrelanger Kritik an der Intransparenz der Moderationspraxis wurden von Facebook 2018 die internen Richtlinien für die Inhaltsmoderation veröffentlicht, mit denen im Wesentlichen festgelegt ist, wie die Gemeinschaftsstandrads für die Inhaltsmoderation auszulegen sind. [31] Trotzdem wird weiterhin kritisiert, dass im Einzelfall vielfach nicht nachvollziehbar ist, aus welchem konkreten Gründen Postings auf Plattformen entfernt wurde oder nicht.

Kritisiert wurden des Weiteren die schlechten Arbeitsbedingungen, ungenügende Ausbildung, starke Belastungen und fehlende psychologische Betreuung für Inhaltsmoderatoren, die zum Teil zentral von Billiglohnländern wie Indien und den Philippinen aus agieren, was dazu beiträgt, dass zu wenig sensibel auf die nationalen Kontexte Rücksicht genommen wird.[32] „Mistakes are inevitable given the sheer volume of requests – and the fact that tech companies simply don't know the context and underlying facts for most real-world disputes that surface as removal requests" (Keller 2015).

Darüber hinaus wird festgestellt, dass die Moderationssysteme insgesamt komplex und undurchsichtig sind, und dass für Nutzer und die Öffentlichkeit nicht nachvollziehbar ist, ob und wie Facebook auf Beschwerden reagiert und ob menschliche Moderatoren mit Beschwerden konfrontiert werden. Denn die menschlichen Moderatoren sind in komplexe technische Moderationssysteme eingebettet, und übernehmen potenziell notwenige Moderationsaufgaben, aber sie dienen vermutlich auch dazu „menschliche Entscheidungen" in Systemen zu suggerieren, in denen bereits ein großer Teil der Entscheidungen technisch getroffen wird (Wagner 2019, 10f.)

---

[30] „Accuracy is also an important issue. Our reviewers work hard to enforce our policies, but many of the judgements require nuance and exceptions. (...) The vast majority of mistakes we make are due to errors enforcing the nuances of our policies rather than disagreements about what those policies should actually be. Today, depending on the type of content, our review teams make the wrong call in more than 1 out of every 10 cases." (Zuckerberg 2018)

[31] Im Jahr 2016 berichtet die Süddeutsche Zeitung erstmals ausführlich darüber, nach welchen Regeln deutsche Facebook-Prüfer gemeldete Beiträge untersuchen. Im Jahr 2017 veröffentlichte der Guardian Ausschnitte aus Facebook's bis dahin geheimen Regelwerk. 2018 veröffentlichte Facebook selbst die Moderationsregeln.

[32] 2018 wurden die Arbeitsbedingungen der Inhaltsmoderatoren auf den Philippinen im Dokumentarfilm "The Cleaners" dargestellt. Zuletzt wurden auch die Arbeitsbedingungen von Inhaltsmoderatoren in US amerikanischen Leihfirmen problematisiert (Newton 2019).

Automatisierte Moderation

In Ergänzung zur menschlichen Moderation kommen auf den Plattformen auch automatisierte Systeme zum Einsatz, die zunehmend auf künstlicher Intelligenz und maschinellem Lernen basieren (siehe z.B. Heldt 2019). Automatisierte Systeme sollen problematische Inhalte erkennen (Vormoderation), diese an menschliche Moderatorinnen und Moderatoren zur Kontrolle weiterleiten, aber mitunter auch automatisch filtern. Mit automatisierten Verfahren können mehr Inhalte rascher und noch vor ihrer Veröffentlichung geprüft werden. Damit steigt die Anzahl der gefilterten Inhalte. Mit dem verstärkten Einsatz automatisierter Systeme kommt es zu einer Verschiebung von reaktiver zu proaktiver Inhaltskontrolle, mit der nicht nur gemeldete, sondern potenziell alle hochgeladenen Inhalte kontrolliert werden. Die Eingriffsintensität nimmt zu.

Die Internet-Plattformen investieren verstärkt in Programme und Technologien zur automatischen Erkennung und Filterung von Inhalten auf Basis von künstlicher Intelligenz und maschinellem Lernen. So betreibt Google das Programm „Perspective" zur Identifikation von Hate Speech. Auch bei Instagram wird an einem Algorithmus zur Erkennung von Hass gearbeitet.[33] Mark Zuckerberg (2018) gibt an, dass automatisierte Systeme bei Facebook bereits 99% aller terroristischen Inhalte und 96% der Nacktbilder, die gegen Community-Richtlinien verstoßen, erkennen, bevor diese gemeldet werden. Für Hate Speech werden hingegen nur 52% der problematischen Beiträge automatisiert erkannt, weil Text und Sprache schwieriger zu beurteilen sind als Bilder.

Der Trend in Richtung automatisierter Systeme zur Inhaltsregulierung gilt als Reaktion auf die rasant steigende Anzahl an Beiträgen, die täglich hochgeladen werden. Aufgrund der höheren Effizienz und Skalierbarkeit werden automatisierte Systeme von den Plattformbetreibern gerne als Lösung propagiert. „Compared to human content reviewers, the advantage and appeal of automated systems is clearly their scalability. Such systems promise to make the content moderation process much easier, quicker and cheaper than would be the case when using human labour" (Gollatz, Beer & Katzenbach 2018, 6). Zudem können automatisierte Systeme dazu beitragen, die quantitativ überlasteten und auch psychisch stark geforderten Moderatorinnen und Moderatoren zu entlasten (ebd., 9f.)

Mitunter werden automatisierte Systeme auch als Reaktion auf den steigenden politischen Druck gesehen, problematische Inhalte umfassend und rasch zu reduzieren. Zuletzt hat etwa die EU-Kommission den Druck auf Plattformen erhöht gegen illegale Inhalte vorzugehen. Das zentrale Instrument ist eine Datenbank, die gemeinsam von Facebook, YouTube, Twitter und Microsoft betrieben wird und an der sich auch Google+ und Instagram beteiligen. „Darin speichern sie digitale Fingerabdrücke, derzeit über 50.000 sogenannte ‚Hashes', von gemeldeten oder automatisch erkannten Inhalten. Der gemeinsam genutzte Upload-Filter unterbindet

[33] Thompson (2017), https://www.wired.com/story/instagram-launches-ai-system-to-blast-nasty-comments/

dann ein erneutes Hochladen“ (Rudl 2018).[34] Nach den Vorstellungen der Kommission sollen sich auch kleine Plattformen an die automatisierten Systeme der großen Betreiber anschließen.

Die automatisierte Inhaltskontrolle ist mit vielfältiger Kritik konfrontiert: Problematisiert wird insbesondere der *beschränkende Effekt für die Redefreiheit* und zum Zugang zu Information aufgrund von „overblocking“ (Gollatz, Beer & Katzenbach 2018; Keller 2015; Dachwitz 2019). Vor allem Menschrechtsorganisationen (z.B. Article 19, 2018) verweisen darauf, dass automatisierte Moderation zu einer Vorzensur führt, mit der das Recht auf Redefreiheit beschränkt wird, und dass automatisierte Methoden ordentliche Verfahren nach rechtstaatlichen Standards verhindern.[35] Beim Einsatz von sogenannten „Upload-Filtern“ sprechen Kritiker von einer engmaschigen privaten Zensurinfrastruktur, die privaten Anbietern die Entscheidung über die Rechtmäßigkeit von Inhalten überlässt (Rudl 2018).

Problematisiert werden vage Definitionen, die Intransparenz hinsichtlich der Implementierung und Funktionsweise der automatisierten Moderationssysteme sowie Unsicherheiten bezüglich der Verantwortlichkeit von und für maschinelle Entscheidungen. Die Löschungen lassen beispielsweise ungeklärt, „ob es um strafbare Äußerungen ging, und verhindern eher eine gesellschaftliche Auseinandersetzung mit dem zugrunde liegenden Hass und der Menschenfeindlichkeit, so einer der Kritikpunkte. Umgekehrt ist für Plattformbetreiber unklar, ob automatisierte Inhaltsmoderation zu einem Verlust der Haftungsprivilegierung führt, die sie genießen, wenn sie als Intermediäre Inhalte lediglich vermitteln und keine Kenntnis über die potenziell unrechtmäßigen Inhalte haben“ (Krempl 2018).

Eine weitere Gefahr bei automatisierten Moderationssystemen sind systematische Verzerrungen (Bias), mangelnde Treffsicherheit und mangelnde Kontextsensibilität. Hohe Erfolgsraten bei der Erkennung problematischer Inhalte gelingen in der Regel nur bei Routineprozessen für die Erkennung von Inhalten mit Standardmustern. Die Übertragung auf andere Fälle und Kontexte ist hingegen schwierig. Speziell Inhalte, deren Legitimität stark vom jeweiligen Kontext abhängt, können von automatisierten Systemen kaum korrekt beurteilt werden. „(T)he more ambiguous and contextual classificatory criteria become, the more difficult it becomes to train algorithms accurately“ (Gollatz, Beer & Katzenbach 2018). Dies gilt beispielsweise für die Einordnung von Satire (vs. Beleidigung), Nacktheit (vs. Kunst) oder Kriegsberichten (vs. terroristische Propaganda). So wurden auf YouTube Kriegsberichte aus Syrien als extremistisch eingestuft und eliminiert, die der Dokumentation von Kriegsverbrechen dienten.

Aufgrund inhärenter Mängel automatisierter Systeme („overblocking“, fehlende Kontextsensitivität, Transparenz, Verantwortlichkeit) wird die Fortsetzung menschlicher Kontrolle der

[34] Siehe Rudl (2018), https://netzpolitik.org/2018/eu-kommission-will-von-plattformen-freiwillige-und-weitreichende-internetzensur/

[35] “Certain decisions to suppress content are in violation of freedom of expression standards, (…) removal of content does not respect requirements of due process of law, (…) majority of decision making processes and practices ... remain opaque“ (Article 19, 2018).

Entscheidungen gefordert sowie die Beachtung demokratischer Prinzipien beim Design und der Anwendung von automatisierter Inhaltsregulierung mittels guter Verständlichkeit und umfassender öffentlicher Beteiligung (Gollatz, Beer & Katzenbach 2018, 8).

Inhaltsregulierung durch Kuratierung

Inhaltsregulierung auf Internet-Plattformen erfolgt nicht nur durch das Löschen von Inhalten, sondern auch durch andere Formen der Inhaltsmoderation. So bleiben manche Inhalte prinzipiell sichtbar, sie werden aber z.B. im Rahmen der Kuratierung weniger prominent dargestellt (De-Priorisierung). Andere Inhalte werden als potentiell problematische Quellen und Inhalte gekennzeichnet und mitunter mit Warnungen oder Zugangsschranken versehen, z.B. zum Jugendschutz. Bei anderen Arten von Inhalten werden gezielt Verlinkungen gesetzt, die auf ergänzende Informationen oder auf Gegenmeinungen verweisen.

Facebook hat in den vergangenen Jahren verschiedene Tools im Rahmen der Kuratierung entwickelt. Dabei geht es meist darum die Transparenz zu erhöhen und die Steuerungsmöglichkeiten der Nutzer zu verbessern. Dazu zählen Funktionen wie „unfollow" oder „see less" für erhöhte Nutzerkontrolle. Transparenzorientierte Formen der Kuratierung wurden auch als Reaktion auf die Desinformationskampagnen eingeführt. So kooperiert z.B. Facebook seit 2016 mit Partnern für Faktenchecks und bietet für ausgewählte Beiträge zusätzliche Kontextinformation, die über einen i-Button erreicht werden (Ash et al. 2019, 15f). Transparenz durch ergänzende Information ist darüber hinaus bei politischer Werbung wichtig. Werbung soll als solche gekennzeichnet und von anderen Inhalten getrennt werden. Nutzer sollen erkennen wer politische Werbung finanziert und über den finanziellen Umfang der politischen Werbung soll Transparenz hergestellt werden. Dies wird insbesondere nach den negativen Erfahrungen mit versteckten Informationskampagnen verstärkt eingefordert.[36]

Solche transparenz-orientierte Ansätze sind weniger eingriffsintensiv als Inhaltsmoderation durch das Löschen von Inhalten. Sie beeinflussen und prägen jedoch ebenfalls die Vermittlung von Information und unsere Wahrnehmung der Welt und verdeutlichen die heikle Rolle der Plattformbetreiber als Regulatoren gesellschaftlicher Kommunikation.

Aktueller Umfang der Inhaltsmoderation

Nach jahrelanger Kritik an der Intransparenz der Inhaltsregulierung wurden von einigen Internet-Plattformen 2018 erstmals Daten zur Inhaltsregulierung auf Basis der Community-Standards in Transparency Reports dokumentiert.[37] Die publizierten Daten verweisen u.a. auf die

[36] Siehe z.B. die Transparenzforderungen der Mozilla Foundation (2019) gegenüber Facebook der sich eine Reihe von NGOs angeschlossen haben.

[37] Z.B. von Facebook: https://transparency.facebook.com/community-standards-enforcement

Arten und den Umfang[38] problematischer Inhalte, gegen die seitens der Plattformen verschiedene Maßnahmen gesetzt werden und über den Anteil problematischer Inhalte, die durch automatisierte Verfahren identifiziert werden konnten.

Darstellung von Kindesmissbrauch: Die Pflicht der Plattformen, die Darstellung von Kindesmissbrauch zu unterbinden, ist politisch und gesellschaftlich unumstritten.[39] Facebook, YouTube und Twitter setzen Software (z.B. Photo DNA) zur automatischen Erkennung bereits gemeldeter Bilder ein, die mit einem digitalen Fingerabdruck versehen werden. Zwischen Juli und September 2018 wurden von Facebook weltweit 8,7 Mio. Inhalte entfernt, die die Nacktdarstellung von Kindern oder deren sexuelle Ausbeutung betrafen. Diese Zahl umfasst auch Fälle, in denen Nutzerinnen und Nutzer „in bester Absicht Nacktfotos ihrer eigenen Kinder teilen".[40] In 99,2% der Fälle wurden Bilder von Facebook identifiziert und gelöscht, noch bevor sie von Nutzern gemeldet wurden.[41]

Terrorismus: Terroristische Inhalte umfassen Material, das von Terrororganisationen produziert wurde, Terrorismus fördert oder verherrlicht und zu Propaganda- und Rekrutierungszwecken eingesetzt wird. Ebenso wie bei Kindesmissbrauch ist die Pflicht diese Inhalte zu löschen weitgehend unumstritten. Facebook und Twitter geben an, über 90% von terroristischen Inhalten durch Algorithmen erkennen zu können.[42] Wie bei der Bekämpfung von Kindesmissbrauchsdarstellung verwendet YouTube eine Datenbank von terroristischen Inhalten mit digitalen Fingerabdrücken, um erneutes Hochladen zu unterbinden.[43]

Urheberrechtsverletzung: Urheberrechtsverstöße stellen vor allem für YouTube eine Herausforderung dar; die Möglichkeit, durch Werbung finanziell zu profitieren, bietet einen Anreiz zum Hochladen urheberrechtsgeschützter Inhalte. Die Plattform setzt das Content ID System ein, das automatisch urheberrechtgeschütztes Material beim Hochladen erkennt. Auch auf anderen Plattformen werden Urheberrechtsverstöße erkannt. Im zweiten Halbjahr 2018 wurden auf Twitter 277.509 Medieninhalte wegen Urheberrechtsverletzungen gesperrt und auf Facebook etwas mehr als 1,7 Millionen Inhalte.

---

[38] Laut Caplan (2018, 24) löschte z.B. YouTube im letzten Jahresviertel 2017 insgesamt 8,2 Millionen von 28 Millionen als problematisch gekennzeichneten Videos. Von den gelöschten Videos wurden 6,5 Millionen Videos durch automatisierte Systeme gekennzeichnet, 1,1 Millionen Videos von „trusted flaggern" und 400.000 Videos von normalen Nutzern.

[39] Gillespie (2018).

[40] Wenn Nutzerinnen und Nutzer auch ohne Missbrauchsabsicht Nacktbilder ihrer Kinder teilen, werden diese von Facebook gelöscht, weil sie „die Möglichkeit des Missbrauchs durch Dritte eröffnen". Facebook (2018). Gemeinschaftsstandards – Nacktdarstellung von Kindern und deren sexuelle Ausbeutung. https://www.facebook.com/communitystandards/child_nudity_sexual_exploitation

[41] Facebook (2018). Facebook Transparency Report. https://transparency.facebook.com/community-standards-enforcement

[42] Twitter (2018). Twitter Rules Enforcement. https://transparency.twitter.com/en/twitter-rules-enforcement.html

[43] Google (2018). YouTube-Community-Richtlinien und ihre Anwendung, Wichtige Richtlinien: Gewalttätiger Extremismus. https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism

Pornographie und Nacktdarstellung: Pornographie zählt zu den legalen, aber gefährlichen Inhalten für Kinder- und Jugendliche und ist nicht auf allen Plattformen untersagt oder unerwünscht. Während Pornographie auf Facebook verboten ist, ist sie auf Twitter erlaubt, wobei Inhalte als sensible Medien gekennzeichnet werden müssen und weiteren Einschränkungen unterliegen.[44] Facebook hat 2018 pro Quartal in 20 bis 30 Millionen Fällen Maßnahmen wegen Nacktheit von Erwachsenen und Pornographie ergriffen und entsprechende Bilder gelöscht.

Cybermobbing und Belästigung: Die Bekämpfung des Cybermobbings dient dem Schutz von Kindern und Erwachsenen vor Belästigung durch einzelne Nachrichten oder mittels koordinierter Kampagnen. In Österreich hat die Veröffentlichung einer obszönen Facebook-Nachricht an die ehemalige Nationalratsabgeordnete Sigrid Maurer 2018 die Debatte über den Umgang mit Belästigung von Frauen im Netz angefacht. Auch koordinierte Kampagnen bergen Sicherheitsrisiken für Opfer, wie etwa die Veröffentlichung persönlicher Daten („doxxing"). Facebook hat 2018 pro Quartal in rund 2 Mio. Fällen Maßnahmen aufgrund von Bullying und Belästigung ergriffen, wobei 85% der Fälle von Nutzern gemeldet werden, ehe sie entfernt werden.

Gewaltdarstellung: Gewaltdarstellung umfasst die Darstellung körperlicher Verletzungen, Gewalttaten sowie Tiermissbrauch. In manchen Fällen hat die mediale Darstellung von Gewalt einen Nachrichtenwert; in manchen kann sie für wichtige Themen sensibilisieren oder sogar als künstlerischer Selbstausdruck verstanden werden. Auf Twitter und Facebook dürfen manche solcher Inhalte eingeschränkt geteilt werden, indem sie als sensible Medien gekennzeichnet und ausgeblendet werden, damit sie erst per Mausklick wieder eingeblendet werden können. Nichtsdestotrotz heben gewalttätige Inhalte das Potenzial, das psychische Wohlbefinden der Nutzer zu beschädigen.[45] Unter anderem nutzen Online-Communitys soziale Medien, um Jugendliche zur Selbstverletzung oder zu Essstörungen zu ermutigen, wie etwa Magersucht bewerbende Fotos auf Instagram.[46] Facebook hat im vergangenen Jahr 1 bis 15 Mio. Inhalte pro Quartal wegen Gewaltdarstellungen gelöscht. Unter 10.000 Postings werden 23 bis 27 wegen Gewaltdarstellung gelöscht.

Hassrede: Hassrede umfasst Angriffe, Gewaltandrohungen oder Aufrufe zu Gewalt gegen Personen oder geschützten Gruppen aufgrund von Eigenschaften wie Geschlecht, Herkunft oder religiöser Zugehörigkeit. Auch Fälle, in denen Personen oder Gruppen „körperlicher Schaden,

[44] Plattformen müssen selbst über den Umgang mit legalen pornographischen oder anzüglichen Inhalten entscheiden, obwohl das Thema auch an Fragen der Persönlichkeitsrechtsverletzung angrenzen kann. Nacktbilder können als Mittel zum Cybermobbing oder zur Erpressung verwendet werden, wenn solche Inhalte ohne Zustimmung der Beteiligten hochgeladen werden (genannt „Racheporno", "nicht-einvernehmliche Pornographie" oder „bildbasierter Missbrauch").

[45] Sie z.B. die Einschätzung von Twitter (2018). Twitter Rules Enforcement. https://transparency.twitter.com/en/twitter-rules-enforcement.html

[46] Burgard-Arp (2018), https://www.zeit.de/campus/2018-09/magersucht-bulimie-pro-ana-zugehoerigkeit-social-media/komplettansicht

Tod oder eine Krankheit"[47] gewünscht wird, können als Hass verstanden werden. Auf Twitter ist Hass schürendes Verhalten die häufigste Ursache für die Sperrung von Konten (Januar-Juni 2018: 285.393 Konten).[48] Facebook ergriff 2018 in 1.5 bis 3 Mio. Fällen Maßnahmen gegen Hasspostings. Automatische Filterung bringt in 52% der Fälle korrekte Ergebnisse. In Österreich stellt Hass gegen Migranten eine ernstzunehmende Herausforderung dar.[49]

Kriminalität und Betrug: Soziale Netze können verwendet werden, um kriminelle Aktivitäten zu unterstützen, zu organisieren, zu verherrlichen oder dazu zu ermutigen, wie zum Beispiel zum Verkauf illegaler Waren und zum illegalen Verkauf regulierter Produkte, zum Betrug oder zur Organisation und Koordinierung krimineller Handlungen.[50] In Österreich wie in vielen Ländern werden soziale Medien genutzt, um Nutzer auf dubiose Webseiten locken[51] und, wie auch in Deutschland bekannt wurde, für hochriskante Finanzglückspiel-Portale zu werben (CFD-Trading).[52]

Einspruchsmöglichkeiten

Inhaltsregulierung ist eine sensible Aufgabe und nicht vor Fehlern sicher. Zudem befinden sich Inhalte immer wieder in Graubereichen, für die Moderationsentscheidungen nicht einfach zu treffen sind. Egal ob Inhalte dann gelöscht werden oder sichtbar bleiben, für Beschwerdeführer, deren Reklamationen zurückgewiesen werden, und für Sprecher, deren Beiträge gesperrt werden, stellt sich die Frage, wie Moderationsentscheidungen beeinsprucht werden können und wie mit Einsprüchen gegen Moderationsentscheidungen umgegangen wird.

Plattformen wurde lange Zeit dafür kritisiert, überhaupt keine formalen Berufungsmöglichkeiten gegen Moderationsentscheidungen anzubieten. Facebook begann erst 2018 mit der Einrichtung eines Berufungsverfahrens. Einschätzungen von Ash (2019, 13) zufolge wird ein relativ hoher Anteil von ca. einem Achtel („perhaps as many as one in eight") an Moderationsentscheidungen beeinsprucht, insbesondere in den Kategorien Hassrede und Bullying. Beschwerden werden bis dato allerdings ebenfalls plattformintern abgewickelt, indem ein beanstandeter Inhalt mit ergänzenden Kontextinformationen einem zweiten Moderator zur Entscheidung vorgelegt wird (Ash et al. 2019, 10). Insgesamt besteht daher der Vorwurf an die

[47] Twitter. Die Twitter Regeln. https://help.twitter.com/de/rules-and-policies/twitter-rules

[48] Twitter (2018). Twitter Rules Enforcement. https://transparency.twitter.com/en/twitter-rules-enforcement.html

[49] Beispielhaft dafür steht das Wiener Neujahrsbaby 2018. Aufgrund des Migrationshintergrunds der Eltern und des Kopftuchs der Mutter posteten etliche Facebook-Nutzer Hasskommentare zum Foto vom ersten Kind, das zum Neujahrstag 2018 in einem Wiener Krankenhaus geboren wurde. Siehe u.a. Wallner, A. (2018), https://diepresse.com/home/meinung/kommentare/5347469/Hass-statt-Honig-fuer-Wiener-Neujahrsbaby-Asel

[50] Facebook (2018), Begünstigung und Propagierung von Verbrechen. https://www.facebook.com/communitystandards/promoting_publicizing_crime

[51] Wannenmacher (2018), https://www.mimikama.at/allgemein/fake-zib/

[52] Pichler (2018), https://derstandard.at/2000072859462/Komm-in-die-Gruppe-Wenn-auf-Whatsapp-der-Reichtum-winkt

Plattformen, dass keine rechtstaatlichen Mindeststandards für diese sensiblen Moderationsverfahren bestehen. „Im Fall der Gemeinschaftsstandards ist Facebook Gesetzgeber, Polizei, Anwalt und Richter zugleich“ (Meineck 2018; vgl. auch Kadri & Klonick 2019). Ein weiterer Kritikpunkt ist die Frage der Begründung von Moderationsentscheidungen. Plattformbetreiber verweisen in der Regel auf Verstöße gegen „Community-Standards“. Solche unkonkreten Begründungen gelten für Beschwerdeführer, deren Inhalte gelöscht werden oder von denen problematische Inhalte gemeldet wurden und deren Ansichten von Inhaltsmoderatoren nicht gefolgt wird, als wenig befriedigend.

Reformüberlegungen der Plattformbetreiber

Aufgrund der weitreichenden Kompetenzen und problematischen Implikationen ihrer Inhaltsregulierung geraten die Plattformen immer stärker in die Kritik und unter Druck von unterschiedlichen Seiten. Dieser Druck („public outrage“) hat auch zu Reformüberlegungen geführt. Exemplarisch dafür stehen Ankündigungen, die Mark Zuckerberg 2018 unter dem Titel „A Blueprint for Content Governance and Enforcement“ veröffentlichte.[53]

Zuckerberg kündigt darin u.a. an, in Zukunft verstärkt auf *künstliche Intelligenz* bei der Identifikation problematischer Inhalte zu setzen und so von einem „reaktiven Ansatz“ auf Basis menschlicher Moderation zu einem „proaktiven Ansatz“ auf Basis technischer Moderation zu wechseln.

Zudem soll auf Basis von künstlicher Intelligenz verstärkt sensationalistischer und provokativer „borderline content“ identifiziert werden, der an der Grenze zu unerlaubten Inhalten liegt und oft besonderes hohe Aufmerksamkeit erhält (z.B. Clickbait, Desinformation, offensive Sprache). Für diese Art von Inhalten soll „die Verbreitungskurve durch die Plattformen verändert werden“,[54] um die Anreize für die Produktion von „borderline content“ zu senken.[55]

In Folge der erwarteten Verbesserungen im Bereich der Identifikation von Inhalten durch künstliche Intelligenz sollen in Zukunft auch die Nutzer selbst mehr Kontrolle über Inhalte bekommen. Wenn Inhalte korrekt identifiziert und beschrieben sind, sollten Nutzer selbst die Entscheidung treffen, welche Arten von Inhalten sie angezeigt bekommen, soweit diese nicht gegen Community-Standards verstoßen. Dafür dürften Überlegungen zu individuell konfigurierbaren Filtern eine Rolle spielen.[56] Diese sollen flexible Standards für Inhaltstypen erlauben, in denen kulturelle Normen und individuelle Präferenzen stark divergieren.

---

[53] Zuckerberg (2018), https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/

[54] Konkreten Informationen wie genau das gemacht werden soll fehlen bislang.

[55] Auch YouTube hat angekündigt stärker gegen borderline content vorzugehen. Siehe YouTube (2019), https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html

[56] Siehe Zuckerberg (2016), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10103508221158471/ und zusammenfassend Hurtz (2017), https://www.sueddeutsche.de/digital/mark-zuckerberg-bald-kann-jeder-facebook-nutzer-entscheiden-wie-viel-nackte-haut-er-sieht-1.3384650

Des Weiteren wurde von Mark Zuckerberg (2019) angekündigt den Schutz der Privatsphäre auf den Plattformen zu erhöhen.[57] In Zukunft soll die Kommunikation auf den Plattformen verstärkt in privaten Gruppen und mittels verschlüsselter Dienste erfolgen. Damit würde auch der Umfang an öffentlich zugänglichen problematischen Inhalten sinken, die maßgeblich zu Kritik, Meldungen und öffentlichem Druck zur Inhaltsregulierung beitragen. Hinter dem neuen Fokus auf Privatsphäre bei Facebook liege demnach u.a. ein Versuch, die steigenden Ansprüche an Content Moderation zu bändigen. „But it's also a solution that will push all that false, conspiratorial, violent, harmful, and hateful content off into information ghettos where journalists, researchers, and watchdogs will have a much more difficult time finding it and calling it out" (Silverman 2019).[58]

Neben solchen technischen Lösungen zur Bewältigung von Herausforderungen der Inhaltsmoderation werden auch Veränderungen in organisatorischen Kontrollstrukturen angekündigt. So soll etwa die Transparenz verbessert werden und eine stärkere Kooperation mit der Forschung erfolgen. Zudem kündigt Facebook die Einrichtung eines unabhängigen Gremiums an, das über einen Teil der Einsprüche entscheidet, und dessen Entscheidungen transparent und verbindlich sein sollen (Facebook 2019a, 2019b; Zuckerberg 2018). Ein „Oversight Board for Content Decisions" soll sich aus etwa 40 internationalen, unabhängigen Expertinnen und Experten zusammensetzen und besonders umstrittene und kritische Moderationsentscheidungen überprüfen und bei Bedarf korrigieren.[59] Dadurch sollen Moderationsteams von der alleinigen Verantwortung für inhaltliche Entscheidungen entlastet werden, Kontrolle und Verantwortlichkeit sollen verbessert werden, und Entscheidungen sollen nicht nach kommerziellen Interessen getroffen werden.

Die Überlegungen zur Delegation von Berufungsverfahren hinsichtlich der Entscheidungen im Rahmen der Inhaltsmoderation an ein unabhängiges Gremium ist nicht zuletzt Ausdruck der Einsicht, dass den Plattformen und v.a. den individuellen Inhaltsmoderatoren zu viel alleinige Verantwortung aufgebürdet wird und diese Verantwortung besser verteilt werden muss. Mark Zuckerberg (2018) gibt zu bedenken: „I've increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own." Allerdings werden die Vorschläge für die institutionelle Ausgestaltung des Oversight Boards ebenfalls kritisiert. Zum Beispiel die Beschränkung des Mandats auf Inhaltsregulierung, die fehlende Verbindlichkeit von Entscheidungen und die direkte Bindung an Facebook wodurch die Unabhängigkeit bezweifelt wird (z.B. Brodnig 2019; Spielkamp 2018). Deshalb erscheinen die Vorschläge für die institutionelle Ausgestaltung verbesserbar, indem z.B. externe Berufung und Aufsicht und damit eine stärkere Verteilung von Verantwortlichkeit ermöglicht werden.

---

[57] Siehe Zuckerberg (2019), A Privacy-Focused Vision for Social Networking, https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/

[58] Silverman, C. (2019), Buzzfeed Newsletter, 14.03.2019.

[59] Facebook (2019a), An Oversight Board for Content Decisions (Draft Charter), https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-2.pdf

3.2 Rechtliche Regulierung von Internet-Plattformen und Inhalten

Neben der Inhaltsregulierung durch Plattformbetreiber auf Basis von Community-Standards bestehen auch rechtliche Vorgaben für Internet-Plattformen und die darüber bereitgestellten Inhalte. Maßgebend sind der europa- und verfassungsrechtliche Rahmen, insbesondere der Schutz der Meinungs- und Medienfreiheit, die Regelungen zur Verantwortlichkeit der Internetdienste-Anbieter für Internet-Inhalte und die auf Internet-Inhalte anwendbaren straf- und zivilrechtlichen Regelungen, die Inhalte wie Kindesmissbrauchsdarstellungen, Beleidigungen, Verhetzung und Cyber-Mobbing unter Strafe stellen.

Grundrechtlicher Rahmen

Zentral für Kommunikation im Internet und auf Internet-Plattformen sind die Meinungs-, Informations- und Medienfreiheit als fundamentale grundrechtliche Normen. Diese Grundrechte schützen die Nutzung des Internet – und zwar im Hinblick auf den Gebrauch der damit verbundenen technischen Infrastruktur, den Zugang zum Internet und die freie Kommunikation unter Nutzung des Internet. Gewährleistet sind die entsprechenden Grundrechte in Bestimmungen des nationalen Verfassungsrechts sowie in Artikel 10 der Europäischen Menschenrechtskonvention (EMRK), die in Österreich in Verfassungsrang steht (vgl. Berka & Trappel 2019). Kommunikation im Internet und auf Internet-Plattformen ist dabei in erster Linie durch das *Grundrecht der Meinungsfreiheit* gem. Art 10 Abs. 1 EMRK geschützt. Einschränkungen der Freiheit zur Mitteilung und zum Empfang von Informationen müssen gemäß Art 10 Abs. 2 EMRK gesetzlich vorgesehen sein, einem bestimmten öffentlichen Interesse oder den Rechten anderer dienen und sie müssen verhältnismäßig sein. Unzulässig und verfassungswidrig sind staatliche Zensurmaßnahmen. Darüber hinaus lassen sich aus einzelnen Grundrechten auch positive Handlungspflichten für den Staat ableiten, der deshalb „mit den ihm zur Verfügung stehenden Mitteln dafür zu sorgen hat, dass nicht im Wege des Internet von dritter, nichtstaatlicher Seite Grundfreiheiten und Menschenrechte verletzt werden“ (Berka & Trappel 2019, 5).

Providerhaftung

Zu den zentralen Regelungen für Internet-Plattformen zählen Vorschriften zum elektronischen Geschäftsverkehr (E-Commerce-Richtlinie 2000/31/EG), denn diese regeln u.a. die Verantwortlichkeit der Internetdienste-Anbieter für Internet-Inhalte. Mit der Richtlinie wurden bestimmte Arten von Internet-Plattformen einheitlich von der Haftung für rechtswidrige Inhalte und Aktivitäten im Netz befreit, die sich ihrer Kenntnis und Kontrolle entziehen. Damit wurde ein technologieunabhängiges rechtliches Umfeld geschaffen, das die Weiterentwicklung von Plattformen erheblich vereinfacht hat.

Basierend auf der europäischen E-Commerce-Richtlinie sieht das österreichische E-Commerce-Gesetz (ECG) Haftungsbefreiungen für Access Provider und Host Provider vor. Ein Host Provider, der Speicherplatz für fremde Inhalte anbietet (z.B. soziale Netzwerke), ist von

der Haftung befreit, sofern er von der Rechtswidrigkeit keine tatsächliche Kenntnis hat und sich in Bezug auf Schadenersatzansprüche auch keiner Tatsachen oder Umstände bewusst ist, aus denen eine rechtswidrige Tätigkeit oder Information offensichtlich wird, oder, sobald er diese Kenntnis oder dieses Bewusstsein erhalten hat, unverzüglich tätig wird, um die Information zu entfernen oder den Zugang zu ihr zu sperren (siehe § 16 ECG; vgl. Zankl 2016, 192f.). Dienste-Anbieter sind gemäß § 18 ECG *nicht verpflichtet*, die von ihnen gespeicherten, übermittelten oder zugänglich gemachten Informationen allgemein zu überwachen oder von sich aus nach Umständen zu forschen, die auf rechtswidrige Tätigkeiten hinweisen. § 18 ECG normiert jedoch auch besondere „Mitwirkungspflichten". So müssen Provider beispielsweise auf Anordnung von Gerichten oder Verwaltungsbehörden Informationen zur Ausforschung von Nutzern übermitteln. Unter gewissen Voraussetzungen sind auch gegenüber Privatpersonen auf deren Verlangen Informationen offenzulegen. [60]

Mit den Haftungsregeln der E-Commerce-Richtlinie wurde „notice and takedown" zum zentralen Regulierungsansatz für gesetzeswidrige Inhalte auf Internet-Plattformen (vgl. Fiedler 2019). Kritisiert wird an der Haftungsprivilegierung unter anderem, dass Internet-Plattformen gegenüber traditionellen Medien bevorzugt werden, für die eine Herausgeberhaftung und strenge inhaltliche Vorgaben gelten, z.B. hinsichtlich Werbung, Jugendschutz und Quotenregelungen, etc. Mittlerweile wird häufig argumentiert, dass die asymmetrische Regulierung zu unfairen Marktbedingungen und ungleichen Schutzstandards für das Publikum führt, worauf z.B. auch Regulierungsbehörden wie Ofcom hinweisen: „As a regulator, we are required to keep audiences safe and protected – irrespective of the screen they watch or the device they hold. Without even knowing it, viewers are watching the same content governed by different regulation in different places or none at all." Die Vorsitzende von Ofcom, Sharon White, beschreibt dies als "standards lottery".[61] Zudem wird argumentiert, dass die Haftungsprivilegierung generell veraltet ist, weil Internet-Plattformen heute kaum mehr als neutrale Anbieter von Hostingdiensten auftreten, sondern in ihre Rolle als Intermediäre die Kuratierung der Inhalte organisieren. Entsprechend wird eine Reform der E-Commerce-Richtlinie vorgeschlagen, um den ursprünglichen Zweck der Richtlinie zu erfüllen (z.B. House of Lords 2019, 52).

Medienrechtliche Regelungen

Internet-Plattformen, deren Inhalte von den Nutzern gestaltet werden (z.B. YouTube), unterliegen nach der geltenden Rechtslage in Österreich nicht dem Medienrecht. Jedoch gelten für bestimmte Angebote, die über Internet-Plattformen verbreitet werden, medienrechtliche Bestimmungen nach dem Mediengesetz (MedienG) und dem Gesetz über audiovisuelle Mediendienste (AMD-G). So unterliegen YouTube-Kanäle, die gedanklichen Inhalt an einen größeren Personenkreis verbreiten, dem Mediengesetz. Zudem kann ein Betreiber eines YouTube-Kanals als Mediendienste-Anbieter anzusehen sein, wenn seine verbreiteten Inhalte mit einer

[60] Zusammengefasst nach Rechteinfach.at (2015).

[61] Waterson (2018), https://www.theguardian.com/media/2018/sep/17/ofcom-to-push-for-regulation-of-social-media-platforms

Fernsehsendung vergleichbar sind. Für Mediendienste-Anbieter bestehen Aufzeichnungs- und Kennzeichnungspflichten, Werberegulierungen, Jugendschutzbestimmungen sowie grundlegende inhaltliche Schranken wie die Achtung der Menschenwürde und ein Verbot der Aufreizung zu Hass (vgl. Berka & Trappel 2019, 13).

Keine medienrechtliche Regulierung gibt es gegenwärtig für Netzwerke und Plattformen, wenn deren Betreiber keine redaktionelle Verantwortlichkeit wahrnehmen, also abgesehen von einer punktuellen Ex-post-Kontrolle keinen Einfluss auf die verbreiteten Inhalte und die Gestaltung des Gesamtangebots nehmen (Berka & Trappel 2019, 14). In vielen Fällen wird allerdings der jeweilige Nutzer, der als Poster, Blogger oder als Gestalter von Videos Inhalte auf Plattformen hochlädt, als Medieninhaber im Sinne des Mediengesetzes anzusehen sein, wenn er für die inhaltliche Gestaltung verantwortlich ist und die Verbreitung selbst vornimmt oder veranlasst. Für Profile auf Facebook ist deshalb der Profilbetreiber medienrechtlich verantwortlich. Diese Personen trifft dann eine Haftung nach dem Mediengesetz sowie die Verantwortlichkeit für die Einhaltung der sonstigen medienrechtlichen Pflichten. Die Haftung des Medieninhabers ist eingeschränkt, wenn es sich um Inhalte Dritter handelt, die auf einer Website bereitgestellt werden und der Medieninhaber dieser Seite nicht „die gebotene Sorgfalt außer Acht gelassen hat". Diese Regelung zielt in erster Linie auf Postings durch dritte Personen, wie sie etwa in der Form von Online-Leserbriefen oder Eintragungen auf Online-Diskussionsforen vorgenommen werden. Unter dem Gesichtspunkt der gesetzlich dem Medieninhaber auferlegten Sorgfaltspflichten ist keine Vorabkontrolle von User-Beiträgen gefordert, sondern im Regelfall ebenfalls nur die unverzügliche Entfernung des inkriminierten Inhalts (vgl. Berka & Trappel 2019, 14f.).

Staatliche Inhaltsregulierung

Rechtmäßigkeitsgrenzen für die Kommunikation auf Internet-Plattformen werden durch straf- und zivilrechtliche Regelungen im öffentlichen Interesse oder zum Schutze der Rechte anderer bestimmt. Die strafrechtlichen Grenzen (siehe Berka & Trappel 2019, 20f.) umfassen in Österreich u.a. das Verbot von Kindesmissbrauchsdarstellungen und harter Pornographie, die Aufforderung zu terroristischen Straftaten oder ihre Gutheißung, Religionsstörung oder das Verbot der Leugnung, Verharmlosung oder Rechtfertigung der nationalsozialistischen Verbrechen gegen die Menschlichkeit („Holocaust-Leugnung").

Von hoher praktischer Relevanz für die Internetkommunikation sind zudem der Schutz von Persönlichkeitsrechten durch den strafrechtlichen Beleidigungsschutz (Üble Nachrede, Beleidigung, Verleumdung) sowie Straftatbestände zum Schutz vor Verletzungen der Menschenwürde (Cyberstalking, Cyber-Mobbing, Verhetzung). Hinzu kommen die zivilrechtlichen Regelungen und Entschädigungstatbestände zum Schutz der Ehre und Privatsphäre (Beleidigung, Kreditschädigung, Recht am eigenen Bild). Die Verbreitung von falschen Gerüchten und manipulierten Nachrichten im Netz verstößt nur ausnahmsweise gegen bestehende

Rechtsvorschriften – etwa wenn bei einer Wahl oder Volksabstimmung falsche Nachrichten verbreitet werden.[62]

Berka und Trappel (2019) betonen, dass es in allen diesen Zusammenhängen ständige Judikatur ist, „dass die einschlägigen Tatbestände im Lichte der Grundfreiheiten und Menschenrechte auszulegen und anzuwenden sind, was in der Regel auf eine an den betroffenen Grundrechten orientierte Interessenabwägung zwischen der Meinungsfreiheit und entgegenstehenden Rechten hinausläuft" (ebd., 21). Strafrechtliche Grenzen der Kommunikation mit den Mitteln des Internet sind nur verfassungsmäßig, wenn die Pönalisierung von bestimmten Äußerungen oder Darstellungen zum Schutz überragender Gemeinschaftsinteressen zwingend erforderlich ist (ebd., 20). Beim zivilrechtlichen Persönlichkeitsschutz erfolgt eine Interessenabwägung zwischen der Meinungsfreiheit und dem Recht auf Achtung der Privatsphäre (ebd., 21). Auch bei der Anwendung von Tatbeständen wie Schmähkritik, Verspottung und Verhetzung muss das Grundrecht der Meinungs- und Medienfreiheit respektiert werden, sodass weder eine polemische oder satirische Auseinandersetzung sanktioniert noch die kritische oder auch provokative Auseinandersetzung mit geschichtlichen Ereignissen unterbunden werden darf. Die tatbestandsmäßigen Folgen und Konsequenzen einer strafbaren Verhetzung müssen sorgfältig geprüft werden, um eine Kriminalisierung der freien Meinungsäußerung zu verhindern und zugleich Angriffen gegen die menschliche Würde entgegenzutreten, wobei solche Grenzziehungen nicht leicht sind (ebd., 24). Der Europäische Gerichtshof für Menschenrechte hat die wichtige Unterscheidung zwischen Inhalten gemacht, die „den Staat oder eine Gruppe der Bevölkerung beleidigen, schockieren oder verstören", und Inhalten, die eine tatsächliche und ernsthafte Aufstachelung zu Gewalt und Hass darstellen. Der Gerichtshof hat klargestellt, dass Staaten Letzteres verbieten und unter Strafe stellen dürfen (Europäische Kommission 2016d).

Laut Einschätzung von Berka und Trappel (2019, 92) nehmen die österreichischen Gerichte bei der Anwendung von Tatbeständen mit Bezug zur Inhaltsregulierung im Internet Bedacht auf die Gewährleistung der Meinungsfreiheit, „sodass überschießende und unverhältnismäßige Einschränkungen dieser Freiheit grundsätzlich nicht zu befürchten sind". Auch nach der Meinung der von Berka und Trappel befragten Akteure der Zivilgesellschaft gibt die Judikatur der österreichischen Gerichte keinen Anlass zu Beschwerden. Die geltenden Strafrechtsgrenzen werden als prinzipiell ausreichend angesehen (ebd.).

Jedoch werden von Expertinnen und Experten sowie von Betroffenen immer wieder *Probleme bei der faktischen Rechtsdurchsetzung* ins Treffen geführt. Solche Probleme der Rechtsdurchsetzung zeigen sich u.a. im Zusammenhang mit Beleidigungen und anderen Eingriffen in die Persönlichkeitsrechte insbesondere im Hinblick auf die Extraterritorialität und Grenzenlosigkeit des Internet und die nicht unerheblichen Kosten der Rechtsverfolgung" (Berka & Trappel

[62] Ein anderer einschlägiger Tatbestand, der die Verbreitung falscher, beunruhigender Gerüchte unter der Voraussetzung unter Strafe gestellt hatte, dass ein größerer Personenkreis beunruhigt und die öffentliche Ordnung gestört wurde, wurde im Jahre 2015 aufgehoben (Berka & Trappel 2019, 25).

2019, 92). „Oft entziehen sich die Host-Provider bei der Löschung von rechtswidrigen Inhalten und bei der Ausforschung von Tätern jeder Verantwortung, indem sie aufwändige Meldesysteme vorschieben und sich auf eigenmächtige Gemeinschaftsstandards berufen" (Windhager & Forgó 2017). Facebook, so die Kritik, lösche Inhalte selbst dann nicht, „wenn man einen gerichtlichen Unterlassungstitel vorlegt, der den postenden Nutzer zur Unterlassung verpflichtet. Auch in solchen Fällen kann man von Facebook die standardisierte Rückmeldung erhalten, dass ein Verstoß gegen die Community Standards nicht ersichtlich ist, obwohl der Rechtsverstoß deutlicher gar nicht mehr belegt werden kann" (Stadler 2016). In Österreich sorgte z.B. 2016 ein Video für Aufsehen, auf dem mehrere Jugendliche ein Mädchen misshandelten. Das Video wurde auf Facebook über 4,5 Mio. Mal angesehen und trotz Aufforderung der zuständigen Staatsanwaltschaft von Facebook nicht zeitnah gelöscht. Exemplarisch für Problemstellungen der Rechtsdurchsetzung steht in Österreich auch der Fall Eva Glawischnig die 2016 in einem Posting auf Facebook beleidigt wurde und eine einstweilige Verfügung zur Entfernung des Postings erwirkte. Im Rahmen eines Verfahrens vor dem Europäischen Gerichtshof (EuGH) soll 2019 geklärt werden, ob eine einstweilige Verfügung gegen Facebook nur im jeweiligen Land oder weltweit gelte und ob Facebook nur wortgleiche oder auch „sinngleiche" Beiträge aktiv suchen und entfernen muss.[63] Zudem bestehen Rechtsunsicherheiten und Rechtsschutzdefizite für Internet Service Provider bei Maßnahmen wie Netzsperren zur Durchsetzung des Urheberschutzes ohne richterliche Kontrolle (Trappel & Berka 2019).[64]

Die Bedeutung der Plattformbetreiber für die gesellschaftliche Kommunikation hat zu *anhaltenden Diskussionen* über eine Verschärfung der Haftung oder zur Erweiterung der Verpflichtung zur Löschung rechtswidriger oder sonstwie anstößiger Inhalte geführt. So haben im Herbst 2018 in Österreich sexistische, frauenverachtende und abwertende Kommentare an Sigrid Maurer und Agrarministerin Elisabeth Köstinger das Thema „Hass im Netz" wieder in die Schlagzeilen gebracht. Die österreichische Bundesregierung veranstaltete deshalb im November 2018 einen Gipfel gegen Frauenhass im Netz.[65] Als Ergebnis dieser Besprechung stellte die österreichische Regierung Überlegungen zu einem „digitalen Vermummungsverbot" an. Sie will vorschreiben, dass Online-Plattformen Nutzer bei Gesetzesverstößen namentlich identifizieren. Im Frühjahr 2019 wurde ein dementsprechendes „Gesetz für Sorgfalt und Verantwortung im Netz" zur Begutachtung vorgelegt. An diesem Vorschlag wird u.a. kritisiert, dass damit eine Einschränkung der Anonymität im Internet und Gefahren für Grundrechte, insbe-

[63] Der Standard (2019), https://derstandard.at/2000097895207/Hasspostings-loeschen-Glawischnig-gegen-Facebook-beschaeftigt-EuGH

[64] Um illegale Inhalte im Netz durch Löschung zu beseitigen, stehen zwar die medien- rechtlichen Sicherungsmaßnahmen zur Verfügung, soweit es sich um Online-Medien handelt, die unter das Mediengesetz fallen. Es fehlen allerdings funktionale, den Besonderheiten des Internet Rechnung tragende Äquivalente für die klassischen medienrechtlichen Instrumente der Einziehung und Beschlagnahme. Solange Netzsperren ohne ausreichende gerichtliche Kontrolle nur durch die privaten Akteure, also vor allem die Provider, verfügt werden, besteht die Gefahr des „Overblocking" und wird dem Gebot eines rechtsstaatlichen Verfahrens nicht ausreichend Rechnung getragen" (Berka & Trappel 2019, 91).

[65] Der Standard (2018), https://derstandard.at/2000090200113/Regierung-organisiert-Gipfel-gegen-Hass-im-Netz

sondere für Meinungsfreiheit, Privatsphäre, Gleichbehandlung einhergehen (vgl. Forgó, Johnson & Polychronopouls 2019). Außerdem wird darauf hingewiesen, dass Hassbotschaften schon heute zumeist unter Klarnamen gepostet werden, dass Formen der Ausweispflicht auf den internationalen Plattformen nur schwer umsetzbar sind und seitens der Medien zu einem Verzicht auf Forenangebote, zu einem Overblocking und somit zu einer Selbstzensur führen können. In einem Fernsehinterview im Anschluss an den Gipfel hat der zuständige Bundesminister Gernot Blümel Überlegungen zu einem Netzwerkdurchsetzungsgesetz für Österreich geäußert.[66] In Bezug auf Fragen zu Providerhaftung und Löschverpflichtung erwartete die österreichische Politik sich hingegen „primär eine Lösung auf europäischer Ebene, wobei die Europäische Kommission in erster Linie auf eine entsprechende Selbstverpflichtung der Betreiber von Plattformen, allenfalls auch in Verbindung mit Verfahren der Co-Regelung zielt" (Berka & Trappel 2019, 15).[67]

3.3 Regulierung und Reforminitiativen auf EU-Ebene

Die Europäische Union entwickelt ihre Politik zum Thema Internet-Plattformen u.a. in den Politikfeldern digitaler Binnenmarkt, Wettbewerb, audiovisuelle Medien, Urheberschutz sowie Kinder- und Jugendschutz. Zuletzt wurde den Internet-Plattformen in einer Reihe von Verfahren, Mitteilungen, Workshops, Studien und Konsultationen verstärkte Aufmerksamkeit gewidmet,[68] wobei in den Stellungnahmen der Kommission sowohl die Innovationsmöglichkeiten als auch die regulatorischen Herausforderungen Beachtung finden. Das Thema der Inhaltsregulierung auf Internet-Plattformen wird von der Kommission u.a. unter dem Grundsatz „Verantwortungsvolles Management von Online-Plattformen zum Schutz der Grundwerte" behandelt (Europäische Kommission 2016a, 8f.). Den Reformüberlegungen der Kommission liegt die steigende Bedeutung der Internet-Plattformen für den Zugang zu Informationen und Inhalten zugrunde. Diese Rolle bringe „notwendigerweise eine größere Verantwortung mit sich" (Europäische Kommission 2016a, 8).

Ein Blick auf die Politik und die Reforminitiativen zeigt, dass diese Verantwortung von den Plattformen laut Vorstellungen der EU-Kommission einerseits vor allem als *Selbstverantwortung* wahrgenommen werden soll. Dies lässt sich sowohl am Festhalten an der Haftungsprivilegierung ablesen, als auch daran, dass die EU-Kommission für etliche Problembereiche in Bezug auf Internet-Inhalte die Bedeutung einer Selbstregulierung betont und auf selbstregulatorische Lösungen setzt.

---

[66] Neuwal.com (2018), Interview Blümel, ZIB 2, 14. November 2018, https://neuwal.com/transkript/20181113-armin-wolf-gernot-bluemel-zib2.php

[67] Im Frühjahr 2019 sprach Minister Blümel von einer zeitgemäßen Adaption der Haftungsregeln in der eCommerce Richtlinie als Aufgabe der neuen EU-Kommission (vgl. ORF.at 2019)

[68] Siehe v.a. Europäische Kommission (2015), Mitteilung über eine Strategie für einen digitalen Binnenmarkt für Europa (KOM (2015) 192); Europäische Kommission (2016a), Mittelung über Online-Plattformen im digitalen Binnenmarkt. Chancen und Herausforderungen für Europa (KOM (2016) 288).

Andererseits übt die Kommission zunehmenden *regulatorischen Druck* auf die Industrie aus. Für immer mehr Bereiche der Inhaltsregulierung im Internet werden rechtliche Vorgaben entwickelt, die direkt oder im Weg einer Ko-Regulierung umzusetzen sind (Jugendschutz, Werbung, Urheberschutz, Terrorismus). Damit werden Plattformen stärker in die Verantwortung genommen und das Prinzip der Haftungsprivilegierung erodiert. Darüber hinaus erhöht die EU-Kommission den Druck auf Plattformen, die Probleme durch Selbstregulierung zu lösen, und sie entwickelt selbst die entsprechenden Initiativen.

Um Probleme in Bezug auf rechtswidrige und schädliche Inhalte und Aktivitäten im Internet zu beseitigen, verfolgt die EU einen sektorspezifischen Regulierungsansatz. Dieser umfasst u.a. die *Bekämpfung der Verbreitung von jugendgefährdenden Inhalten und Hassreden*[69] auf Videoplattformen. Das Parlament der Europäischen Union hat dazu im Oktober 2018 die Novelle der EU-Richtlinie für audiovisuelle Mediendienste (AVMD) verabschiedet.[70] Diese sieht eine Erweiterung des Anwendungsbereichs der Richtlinie auf Videoplattformdienste vor. Videoplattformanbieter werden verpflichtet, in ihrem Zuständigkeitsbereich angemessene Maßnahmen zu ergreifen, um Minderjährige vor schädlichen Inhalten und alle Bürgerinnen und Bürger vor Aufstachelung zu Gewalt oder Hass zu schützen. „Im Einzelnen vorgesehen sind Mechanismen, mit deren Hilfe die Nutzer schädliche Inhalte melden und anzeigen können, Altersüberprüfungssysteme oder Systeme zur elterlichen Kontrolle. Um sicherzustellen, dass die Maßnahmen zukunftsfähig und wirksam sind, wird die Kommission alle Videoplattformanbieter einladen, im Rahmen einer Allianz für einen besseren Schutz von Minderjährigen im Internet an einem Verhaltenskodex für die Branche mitzuarbeiten. Anstelle einfacher Selbstregulierung werden die für audiovisuelle Medien zuständigen nationalen Regulierungsstellen die Befugnis zur Durchsetzung der Vorschriften erhalten. Dies kann – je nach den nationalen Rechtsvorschriften – auch Geldstrafen nach sich ziehen."[71]

Seit 2018 wurde von der EU-Kommission, nicht zuletzt mit Blick auf die Wahlen zum Europäischen Parlament, auch die Politik im *Kampf gegen Desinformationen* intensiviert.[72] Dabei wurde neben einem Verhaltenskodex[73] für Plattformunternehmen auch ein Aktionsplan gegen Desinformation entwickelt, der im Dezember 2018 vorgestellt wurde.[74] Das Maßnahmenpaket des Aktionsplans umfasst neben der EU-weiten Umsetzung des Verhaltenskodex und Informationsmaßnahmen zur Sensibilisierung der Gesellschaft auch den Aufbau von Analysekapazitäten und eine verstärkte Zusammenarbeit zwischen den Mitgliedstaaten und den EU-Organen.

---

[69] Siehe dazu auch den Rahmenbeschluss 2008/931/JI zur strafrechtlichen Bekämpfung bestimmter Formen und Ausdrucksweisen von Rassismus und Fremdenfeindlichkeit ABl L 328, 55.

[70] Europäische Kommission (2018d), http://europa.eu/rapid/press-release_MEMO-18-4093_en.htm

[71] Europäische Kommission (2016b), http://europa.eu/rapid/press-release_IP-16-1873_de.htm; Europäische Kommission (2016e), http://europa.eu/rapid/press-release_MEMO-16-1895_de.htm

[72] Europäische Kommission (2018a), Mitteilung zur Bekämpfung von Desinformation im Internet (KOM (2018) 236), Europäische Kommission (2018b), Mitteilung Freie und faire Europawahlen gewährleisten (KOM (2018) 637).

[73] https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[74] JOIN (2018) 36, Aktionsplan gegen Desinformation.

Investitionen in digitale Tools, Datenanalysefähigkeiten und Fachpersonal in den EU-Organen und den Mitgliedstaaten sollen zu einer besseren Erkennung, Untersuchung und Enthüllung von Desinformation führen und eine Bewertung der Tragweite und Auswirkungen von Desinformation ermöglichen. Zudem sind eine engere Koordinierung und gemeinsame Abwehrmaßnahmen gegen Desinformation geplant. Dies soll durch die Einrichtung eines Frühwarnsystems erfolgen, um in Echtzeit vor Desinformationskampagnen zu warnen. Auf nationaler Ebene sollen Kontaktstellen benannt werden, um den Informationsaustausch zwischen den Mitgliedstaaten zu verbessern.

Ebenfalls in den Bereich der Inhaltsregulierung fällt die *Politikentwicklung im Bereich Urheberschutz*, insbesondere durch eine Überarbeitung der EU-Richtlinie über das Urheberrecht im digitalen Binnenmarkt. Seit 2014 wird auf EU-Ebene über ein neues Urheberrecht verhandelt. Im Paket zur Urheberrechtsreform setzt sich die Kommission „für eine gerechtere Verteilung der von Online-Plattformen durch die Verbreitung urheberrechtlich geschützter Inhalte erwirtschafteten Einnahmen ein". Leistungen von Medien und Kulturschaffenden sollen auch im Internet adäquat vergütet werden. Dies soll durch eine Reform des Urheberrechts und die Einführung des Leistungsschutzes auf EU-Ebene erreicht werden. Die Haftung für Uploads durch Nutzer soll in Zukunft bei den großen Plattformen liegen. Sie müssen laut Richtlinie effiziente und proportionale Maßnahmen ergreifen, um Urheberrechtsverletzungen zu verhindern. Das Reformvorhaben wurde 2018/2019 im EU-Parlament diskutiert und hat zu hitzigen politischen Debatten und Demonstrationen geführt. Aus Sicht von Netzaktivisten sind Maßnahmen vorgesehen, die die Freiheit des Internet massiv beschränken könnten. Kritisiert wird v.a., dass jeder Upload auf Plattformen kontrolliert werden muss, um Urheberrechtsverletzungen zu verhindern. „Die Entscheidung darüber treffen bei den großen Plattformen in den meisten Fällen automatisierte Systeme. Diese werden bereits bisher für drastische Fehlentscheidungen verantwortlich gemacht. Gegner sprechen von ‚Zensurmaschinen'."[75]

Des Weiteren spielt für die EU-Politik die *Haftungsregelung* eine zentrale Rolle. Die Kommission setzt sich dafür ein, den derzeitigen Haftungsausschluss für Anbieter von Vermittlungsdiensten und damit „eine ausgewogene und vorhersehbare Haftungsregelung für Online-Plattformen" aufrechtzuerhalten (Europäische Kommission 2016a). Dies sei von entscheidender Bedeutung für die weitere Entwicklung der digitalen Wirtschaft in der EU und für die Mobilisierung von Investitionen in Plattform-Ökosysteme. Dahingehend bestehen jedoch bei den Plattformbetreibern Bedenken, dass sie im Zuge der Einführung freiwilliger Maßnahmen im Rahmen der Selbstregulierung ihren Anspruch auf Haftungsausschluss für Anbieter von Vermittlungsdiensten gemäß der E-Commerce-Richtlinie verlieren könnten. Die Kommission hat deshalb angekündigt zu prüfen, ob in Bezug auf die Haftung von Online-Plattformen bei Einführung freiwilliger vertrauensbildender Maßnahmen zur Bekämpfung illegaler Internet-Inhalte weiterer Klärungsbedarf besteht.

[75] Fanta (2018), https://netzpolitik.org/2018/schlag-gegen-die-netzfreiheit-eu-abgeordnete-treffen-vorentscheid-fuer-uploadfilter-und-leistungsschutzrecht/

Generell strebt die Kommission europaweite Lösungen für Fragen der Plattformregulierungen an und warnt vor nationalen Alleingängen (Europäische Kommission 2016a). So könne Europa mit 28 verschiedenen Regelwerken für Online-Plattformen in einem Binnenmarkt nicht in den Genuss sämtlicher Vorteile der Plattformwirtschaft kommen und das Wachstum europäischer Plattform-Startups anregen. Unterschiedliche nationale oder sogar lokale Vorschriften für Online-Plattformen würden zu Rechtsunsicherheit bei Wirtschaftsakteuren führen, die Verfügbarkeit digitaler Dienste einschränken und Verwirrung bei Verbrauchern und Unternehmen stiften. Laut Einschätzung der EU-Kommission sind harmonisierte Vorschriften auf EU-Ebene wie die Datenschutz-Grundverordnung wichtig zur Förderung des Wachstums und der schnellen Ausweitung innovativer Plattformen.

3.4 Selbstregulierung auf europäischer Ebene

Neben den regulatorischen Reaktionen auf Richtlinien-Ebene tritt die Kommission generell verstärkt für die Förderung koordinierter, EU-weiter *Selbstregulierungsmaßnahmen* von Online-Plattformen ein. Explizit nennt die Kommission Probleme wie Aufstachelung zum Terrorismus, sexuellen Missbrauch von Kindern und Hassreden, für die Plattformbetreiber sensibilisiert werden sollen, um freiwillige Maßnahmen zum Schutz der Nutzer vor rechtswidrigen oder schädlichen Inhalten zu treffen. Dabei übt die EU-Kommission zunehmenden *regulatorischen Druck* auf die Industrie aus und sie initiiert, entwickelt und kontrolliert die „Selbstregulierung" der Industrie immer stärker selbst.

Bereits seit langem etabliert sind Selbst- und Ko-Regulierung für Maßnahmen im Bereich des Kinder- und Jugendschutzes, wo eine Vielzahl an Programmen und Selbstregulierungsinitiativen von der EU initiiert, gefördert und koordiniert wird. So haben sich insgesamt 18 Betreiber von sozialen Online-Netzwerken, darunter auch Facebook, Google und MySpace, auf europäischer Ebene 2008 den Safer Social Networking Principles angeschlossen, mit denen Jugendschutzziele verfolgt werden. Seitens der Industrie wird Selbstregulierung zum Schutz von Kindern und Jugendlichen auch von der ICT Coalition for Children Online und von der CEO Coalition unterstützt. Ebenfalls dem Jugendschutz dienen Programme wie Insafe, die Safer Internet Centers in Mitgliedstaaten, Better Internet for Kids Portale oder Inhope, ein Hotline-Netzwerk im Kampf gegen Kindesmissbrauchsdarstellungen. 2017 wurde die Alliance to better protect minors online als Selbstregulierungsinitiative gegründet, um Kinder vor gefährlichen Inhalten, gefährlichem Verhalten und gefährlichen Kontakten zu schützen.

Des Weiteren setzt die EU-Kommission auch bei der Bekämpfung terroristischer Propaganda auf Zusammenarbeit und Selbstregulierung durch die Industrie. 2015 wurde dafür das EU-Internetforum eingerichtet, um den Missbrauch des Internets durch internationale terroristische Gruppierungen zu beenden. Das EU-Internetforum wird als freiwillige Partnerschaft bezeichnet, an der sich EU-Innenminister, die Internetbranche und andere Interessenträger beteiligen, um den Zugang zu terroristischen Inhalten im Internet einzuschränken und wirk-

same Kampagnen mit alternativen und Gegendiskursen zu erarbeiten. Im Kontext der Initiative erfolgte u.a. der Aufbau einer gemeinsamen „Hash-Datenbank" durch große Plattformbetreiber, die dazu dient, terroristische Inhalte in den sozialen Medien besser zu erkennen und ihr Wiederauftauchen auf anderen Plattformen zu verhindern. 2018 wurde der regulatorische Druck der Kommission mit der Vorlage eines Entwurfs für eine Verordnung[76] zur Verhinderung der Verbreitung terroristischer Online-Inhalte erhöht. Mit einem abgestimmten Rechtsrahmen soll der Missbrauch von Hosting-Diensten für die Verbreitung terroristischer Online-Inhalte verhindert werden. Das Beispiel der Terrorismusbekämpfung steht damit auch exemplarisch für eine *Verschärfung der* Regulierung durch einen Schritt von Selbstregulierung zu rechtlicher Regulierung.

Als Reaktion auf das zunehmende Problem von Hass im Internet legte die Europäische Kommission im Mai 2016 zusammen mit vier großen IT-Unternehmen (Facebook, Microsoft, Twitter und YouTube) einen Verhaltenskodex[77] für die Bekämpfung illegaler Hassreden im Internet vor. Der Kodex enthält Verpflichtungen zur Bekämpfung der Verbreitung von illegaler Online-Hetze in Europa. Dazu gehört u.a. die Verpflichtung, die Mehrheit der stichhaltigen Anträge auf Entfernung illegaler Hasskommentare in weniger als 24 Stunden zu prüfen und erforderlichenfalls zu entfernen, sowie die Verpflichtung zur Stärkung von Partnerschaften mit zivilgesellschaftlichen Meldestellen.[78]

Im Oktober 2018 wurde der Code of Practice on Disinformation[79] vorgestellt, der ebenfalls nicht zuletzt auf Initiative der EU-Kommission zustande gekommen war. Damit geben Unternehmen der Internet- und Werbewirtschaft, darunter auch Google, Facebook, Twitter und Mozilla, eine Selbstverpflichtung zum Kampf gegen Desinformation und manipulierende Wahlwerbung ab. Die Herkunft politischer Werbung soll transparent gemacht werden. Quellen, die Falschinformationen verbreiten, sollen mit dem Entzug von Werbeeinnahmen sanktioniert und gefälschte Accounts schneller gesperrt werden. Zudem sollen die Betreiber stärker gegen Bots vorgehen, die Fehlinformationen im Netz verbreiten. Nutzer sollen mehr Möglichkeiten zur Beschwerde erhalten.[80]

In der Beurteilung der Selbstregulierung spricht die Kommission selbst von „vielen lobenswerten Initiativen", die von den IT-Unternehmen ergriffen wurden. So wurden u.a. die Maßnahmen gegen Hass im Netz nach einem Jahr evaluiert und positiv bewertet. Die Kommission

---

[76] Siehe Europäische Kommission (2018e), Vorschlag für eine Verordnung zur Verhinderung der Verbreitung terroristischer Online-Inhalte.

[77] Verhaltenskodex für die Bekämpfung illegaler Hassreden im Internet. http://ec.europa.eu/newsroom/document.cfm?doc_id=42875; https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300

[78] Europäische Kommission (2016c), http://europa.eu/rapid/press-release_IP-16-1937_de.htm

[79] Siehe EU Code of Practice on Disinformation, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454, https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

[80] Europäische Kommission (2018c), http://europa.eu/rapid/press-release_STATEMENT-18-5914_en.htm; ORF.at (2018), https://orf.at/stories/3102259/

attestierte Verbesserungen bei der Bearbeitung von Nutzerbeschwerden und bei der Zusammenarbeit mit der Zivilgesellschaft.[81] Die Bemühungen im Rahmen der Selbstregulierung müssten laut Einschätzung der Kommission fortgeführt und ausgeweitet werden, damit sich ihre positive Wirkung auf die Gesellschaft noch verstärkt.

Der Drang der EU zu Selbstregulierungsinitiativen wird jedoch auch kritisiert. Einerseits sind die Verhaltenskodizes rechtlich nicht bindend und sehen keinen Durchsetzungsmechanismus vor.[82] Andererseits werden mit der Selbstregulierung rechtstaatliche Standards untergraben und zentrale Schranken für staatliche Inhaltsregulierung umgangen. Direkte hoheitliche Regelungen unterliegen der Kontrolle der Gerichte, von denen bislang z.B. rechtliche Vorschriften zu allgemeinen Überwachungspflichten für Plattformanbieter mittels genereller Filtersysteme untersagt wurden.[83] Die Kommission versuche, so die Kritik, diese Vorschrift zu umgehen, indem sie Druck auf die Plattformen ausübt, Inhalte „freiwillig" zu filtern. Freiwillige Maßnahmen privater Anbieter würden der Kontrolle der Gerichte entzogen,[84] und rechtsstaatliche Standards werden unterminiert (vgl. z.B. Heldt 2019; ausführlich auch Keller 2019).

3.5 Ausgewählte Initiativen in anderen europäischen Ländern

Vereinigtes Königreich: Im Vereinigten Königreich werden Fragen zu Internet-Plattformen und Inhaltsregulierung auf parlamentarischer Ebene in einer Reihe von aktuellen Untersuchungsausschüssen und -berichten thematisiert, darunter „Disinformation and fake news"[85], „Regulating in a digital world"[86], „Growing up with the Internet"[87] oder „Algorithms in decision making"[88]. Seitens der Regierung wurde der Reformprozess mit einem Grünbuch zur Internet Safety Strategy (2017)[89] initiiert, eine Digital Charter angekündigt und ein „White Paper on online harms " (HM Government 2019) veröffentlicht. Der Diskussionsprozess zur Zukunft der Internet-Regulierung ist damit voll im Gange. Eine der parlamentarischen Inquiries widmet sich dem Thema „The Internet: to regulate or not to regulate". In etlichen begleitenden

[81] Europäische Kommission (2017), http://europa.eu/rapid/press-release_IP-17-1471_de.htm

[82] Siehe z.B. die Kritik des Sounding Board of the Multistakeholder Forum on Disinformation Online (2018) am Code of Practice on Desinformation.

[83] Sie SABAM vs. Netlog (C-360/10), mit dem der Europäische Gerichtshof entschied, dass ein soziales Netzwerk *"cannot be obliged to install a general filtering system, covering all its users, in order to prevent the unlawful use of musical and audio-visual work"*.

[84] Siehe u.a. Rudl (2018), https://netzpolitik.org/2018/eu-kommission-will-von-plattformen-freiwillige-und-weitreichende-internetzensur/

[85] Siehe House of Commons: Digital, Cultural, Media and Sport Committee (2019).

[86] Siehe House of Lords: Select Committee on Communications (2019).

[87] Siehe House of Lords: Select Committee on Communications (2017).

[88] Siehe House of Commons: Digital, Cultural, Media and Sport Committee (2018).

[89] Siehe https://www.gov.uk/government/consultations/internet-safety-strategy-green-paper

Analysen von NGOs und Think Tanks wird die Einrichtung einer unabhängigen Regulierungsbehörde empfohlen, zuletzt beispielsweise prominent von der LSE Commission on Truth, Trust and Technology (2018) in Form einer Independent Platform Agency.[90]

Frankreich: Als erstes europäisches Land hat Frankreich im Herbst 2018 ein Gesetz gegen die Verbreitung von „Fake News" während eines Wahlkampfs beschlossen.[91] „Die neuen Regeln ermöglichen es u.a., dass sich Kandidatinnen und Kandidaten im Eilverfahren gegen die Verbreitung von Falschinformationen im Internet wehren können. Ein Gericht kann dann etwa einen Suchmaschinenanbieter verpflichten, eine bestimmte Seite nicht mehr in seiner Ergebnisliste anzuzeigen. Bedingung ist, dass die Falschinformationen geeignet sind, die Wahl zu beeinflussen" (ORF.at, 2018). Die französische Opposition übte scharfe Kritik und sprach unter anderem vom „Versuch der Informationskontrolle".

Deutschland: In Deutschland trat 2017 Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz) in Kraft. Das Gesetz zielt darauf ab rechtswidrige Inhalte wie Hasskriminalität, strafbare Falschnachrichten und andere strafbare Inhalte auf den Plattformen sozialer Netzwerke wirksamer zu bekämpfen. Das Gesetz fordert von sozialen Netzwerken die Benennung eines inländischen Zustellungsbevollmächtigten, eine gesetzliche Berichtspflicht für soziale Netzwerke über den Umgang mit Hasskriminalität und anderen strafbaren Inhalten sowie ein wirksames Beschwerdemanagement. Soziale Netzwerke müssen offensichtlich rechtswidrige Inhalte innerhalb von 24 Stunden nach Eingang der Beschwerde löschen oder sperren. Über andere gemeldete Inhalte müssen soziale Netzwerke innerhalb von sieben Tagen nach Eingang der Beschwerde entscheiden. Verstöße gegen diese Pflichten können mit Bußgeldern geahndet werden.

Maßgeblich dafür, welche Inhalte gelöscht werden müssen, sind die deutschen Strafgesetze, und das zuständige Bundesministerium für Justiz betont, dass der Maßstab, was gelöscht werden muss, nicht von den sozialen Netzwerken gesetzt wird" (Bundesjustizamt 2018). Kritiker halten dem jedoch entgegen, dass „die Bundesregierung mit dem NetzDG private Unternehmen zu Richtern über die Presse- und Informationsfreiheit im Netz gemacht habe", weil Dienste-Anbieter selbst entscheiden müssen, ob ein Inhalt rechtswidrig ist, wenn er gemeldet wurde. Kritiker des Gesetzes, wie zum Beispiel Reporter ohne Grenzen, werfen der Regierung vor, dass es durch die Regelung zu einem „overblocking" kommt, weil Dienste-Anbieter potenziell problematische Inhalte sicherheitshalber entfernen, um nicht mit dem Gesetz in Kon-

[90] Siehe LSE (2018) sowie zur Empfehlung für eine unabhängige Regulierungsbehörde z.B. auch Doteveryone (2018), Bunting (2018). Für einen Überblick zu aktuellen Regulierungsinitiativen LSE (2018, 46f.). Der Willensbildungsprozess ist noch nicht abgeschlossen. Gerechnet wird jedoch mit einer Verschärfung der Regulierung durch eine Einführung von Sorgfaltspflichten (duty of care), einem Verhaltenskodex und einer Aufsichtsbehörde, deren Durchsetzungsinstrumente noch Gegenstand von Verhandlungen sind.

[91] Siehe Proposition de loi relative à la lutte contre la manipulation de l'information. http://www.assemblee-nationale.fr/15/ta/tap0190.pdf

flikt zu geraten und belangt zu werden. Sie fordern eine öffentliche Kontrolle des Löschverfahrens durch eine unabhängige Prüfinstanz sicherzustellen, um das Löschen von rechtlich zulässigen Inhalten zu erkennen und „overblocking“ zu verhindern (Krempl 2018).

„Der staatlich verordnete Löschdruck durch das NetzDG hat laut Reporter ohne Grenzen dazu geführt, dass die Betreiber eigene Community Standards heranziehen, um im Zweifel auch legale Inhalte zu beseitigen. Kämen die betroffenen Konzerne ihren gesetzlich festgeschriebenen Säuberungspflichten systematisch nicht nach, drohten hohe Geldstrafen. Löschten sie dagegen problematische Inhalte bereits großzügig auf Basis ihrer eigenen Vorgaben, entgingen sie dieser Gefahr“ (Krempl 2018).

Mittlerweile liegen erste Daten zur Umsetzung der Vorgaben gemäß NetzDG in Deutschland vor. Bei Daten zu Maßnahmen nach dem NetzDG ist zu berücksichtigen, dass Internet-Plattformen Inhaltsmoderation hauptsächlich nach eigenen Regeln vornehmen, die zum Teil mehr verbieten als gesetzeswidrige Inhalte. Daten zu Löschungen nach dem NetzDG liegen für die erste Jahreshälfte 2018 vor. [92] Bei Google gingen für YouTube im ersten Halbjahr 2018 Beschwerden über 213.330 Inhalte mit Bezug auf das NetzDG ein, von denen der Konzern 27% unter Verweis auf Strafbestände oder die Community-Richtlinien des Videoportals entfernte. YouTube prüft Meldungen sowohl nach den Vorgaben des NetzDG als auch den eigenen Standards und entscheidet dann, einen Inhalt global zu löschen oder für Zugriffe aus Deutschland zu blockieren. Twitter meldete 264.818 Beschwerden und ging gegen 28.645 (11%) der vorgebrachten Inhalte vor. Bei Facebook gingen in den ersten sechs Monaten 886 Beschwerden nach dem Netzwerkdurchsetzungsgesetz ein, wobei Nutzer 1.704 Inhalte beanstandeten. Facebook löschte 362 der gemeldeten Inhalte (21%). Allerdings ist der Meldevorgang bei Facebook komplizierter als bei der Konkurrenz, und Facebook ergreift Maßnahmen gegen Inhalte vor allem auf Basis der Gemeinschaftsstandards, die auch Nacktbilder, Spam und Falschnachrichten untersagen. Wie viele Beiträge Facebook aufgrund eigener Standards ausblendete, wird nicht nach Ländern ausgewiesen. Es dürften jedoch deutlich mehr sein. Die Transparenz und Nachvollziehbarkeit der Art, wie Facebook seine eigenen Regeln durchsetzt, ist allerdings nur äußerst begrenzt (vgl. Krempl 2018).

Insgesamt ist die Bilanz der Effekte des Netzwerkdurchsetzungsgesetzes gemischt. Einerseits bildet das Gesetz einen ersten Versuch die Rechtsdurchsetzung auf nationaler Ebene zu verbessern. Andererseits besteht Kritik, z.B. aufgrund von möglichem overblocking und Divergenzen bei der Umsetzung der Berichtspflichten der Plattformen, die bei einer Evaluierung des Gesetzes berücksichtigt werden sollte. Zudem ist zu bedenken, dass mit dem NetzDG zwar ein hartes Regulierungsinstrument mit starken Sanktionsmöglichkeiten zur Verfügung steht. Daneben besteht nun jedoch wenig Platz, zum Umgang mit problematischen Inhalten mit weicheren Regulierungsinstrumenten und durch kooperative Strategien.

[92] Krempl (2018), https://www.golem.de/news/netzdg-kritiker-erste-loeschberichte-bestaetigen-gefahr-von-overblocking-1807-135758.html

3.6 Fazit zum Status quo der aktuellen Regulierungspraxis

Zur Erfassung des Status quo der aktuellen Praxis der Inhaltsregulierung sind sowohl die Maßnahmen zu berücksichtigen, die als Selbstregulierung von Internet-Plattformen durch Inhaltsmoderation gesetzt werden, als auch die rechtlichen Vorgaben in Österreich und der EU, mit denen der staatliche Regulierungsrahmen für die Inhaltsregulierung aufgespannt wird. Stand und Kritik der aktuellen Regulierungslage lassen sich wie folgt zusammenfassen.

Im Zentrum der Inhaltsregulierung stehen die Plattformen selbst. Sie legen Gemeinschaftsrichtlinien fest, definieren damit unerwünschte Inhalte und implementieren komplexe Systeme, um Kommunikation auf den Plattformen zu kontrollieren, zu filtern oder im Rahmen der Kuratierung weniger prominent darzustellen. Diese Filterung und Kuratierung wird als Inhaltsmoderation bezeichnet und erfolgt mittels einer Kombination aus Meldesystemen, automatisierten Moderationsverfahren, menschlichen Kontrollen, automatisierter Kuratierung und transparenz-fördernder Kennzeichnung.

Die bestehenden Systeme sind grundsätzlich legitim, denn es liegt im Ermessen der Plattformen zu bestimmen, welche Arten von Inhalten sie tolerieren und wie sie gegen inakzeptable Inhalte vorgehen, solange die Inhalte nicht gegen geltendes Recht verstoßen. Trotzdem wird die Inhaltsmoderation aus vielfältigen Gründen kritisiert:

Die Gemeinschaftsrichtlinien widerspiegeln US-amerikanische Wertvorstellungen und implementieren diese weltweit ohne Rücksicht auf nationale Bedürfnisse. Sie werden als zu unpräzise und zu überschießend kritisiert, sodass sie die freie Meinungsäußerung beschränken.

Die menschliche Inhaltsmoderation ist ressourcenintensiv und wird aufgrund von Intransparenz und Ungenauigkeit problematisiert. Kritisiert werden zudem die schlechten Arbeitsbedingungen, hohe Belastungen und fehlendes Verständnis für nationale Kontextbedingungen der Inhaltsmoderatorinnen und -moderatoren, vor allem in Billiglohnländern.

Die Internet-Plattformen verlassen sich nicht nur auf Meldungen problematischer Inhalte (notice and takedown), sondern investieren auch verstärkt in Programme und Technologien zur automatischen Erkennung und Löschung von Inhalten auf Basis von künstlicher Intelligenz. Dadurch kommt es zu einer Verschiebung von reaktiver zu proaktiver Inhaltskontrolle, mit der nicht nur gemeldete, sondern potenziell alle hochgeladenen Inhalte kontrolliert werden. Diese automatisierte Kontrolle wird kritisiert, weil sie zu einem „overblocking“ und zur Beschränkung der Redefreiheit führen kann. Als weitere Gefahren bei automatisierten Moderationssystemen gelten systematische Verzerrungen (Bias), mangelnde Treffsicherheit und mangelnde Kontextsensibilität, z.B. bei der Beurteilung von Satire oder Kriegsberichten und das Verhindern von gesellschaftlichen Diskursen über die Ursachen für unerwünschte Inhalte. Es bestehen Herausforderungen für den Umgang mit Inhalten, die nach individueller Überprüfung verlangen und sich nicht für automatisierte Moderation eignen.

Zudem wird ein Mangel an rechtsstaatlichen Standards problematisiert, denn trotz der erheblichen Fehlerwahrscheinlichkeit bei automatisierten und manuellen Kontrollen fehlen bislang auf Plattformen effektive Einspruchsmöglichkeiten, unabhängige Berufungsverfahren und nachvollziehbar begründete Moderationsentscheidungen.

Reformüberlegungen bei Facebook zielen darauf ab, verstärkt in automatisierte Verfahren zur Identifikation problematischer Inhalte zu investieren, diese weiterzuentwickeln und mit Hilfe technischer Lösungen Anreize zu reduzieren, Inhalte zu veröffentlichen, die an der Grenze zu Illegalität liegen. Facebook hat angekündigt Nutzern mehr Kontrolle über Inhalte zu geben indem Nutzer verstärkt selbst und technikgestützt bestimmen, welche Arten von Inhalten sie akzeptieren und ablehnen. Für Beschwerden, Aufsicht und Kontrolle werden Überlegungen zu einem Berufungsverfahren, einem unabhängigen Aufsichtsgremium (Oversight Board for Content Decisions), mehr Transparenz und verstärkter Kooperation mit der Forschung und Regulierung angestellt.

Neben der Inhaltsregulierung durch Plattformbetreiber auf Basis von Community-Standards bestehen auch rechtliche Vorgaben für Internet-Plattformen und die darüber bereitgestellten Inhalte. Maßgebend sind der europa- und verfassungsrechtliche Rahmen, insbesondere der Schutz der Meinungs- und Medienfreiheit gemäß Europäischer Menschenrechtskonvention (EMRK), sowie die Regelungen zur Verantwortlichkeit der Internetdienste-Anbieter für Internet-Inhalte insbesondere das Haftungsprivileg auf Basis der E-Commerce-Richtlinie. Gleichzeitig bestehen die auf Internet-Inhalte anwendbaren straf- und zivilrechtlichen nationalen Regelungen, die Inhalte wie Kindesmissbrauchsdarstellungen, Beleidigungen, Verhetzung und Cyber-Mobbing unter Strafe stellen.

Expertinnen und Experten (vgl. Berka & Trappel 2019, 90f.) halten die in Österreich geltenden Strafrechtsgrenzen für Internet-Inhalte für prinzipiell ausreichend. Die österreichischen Gerichte nehmen bei der Anwendung von Tatbeständen mit Bezug zur Inhaltsregulierung im Internet Bedacht auf die Gewährleistung der Meinungsfreiheit, sodass überschießende und unverhältnismäßige Einschränkungen dieser Freiheit grundsätzlich nicht zu befürchten sind. Die Praxis der österreichischen Judikatur zeigt einen guten Ausgleich der verschiedenen Grundrechtsansprüche. Zu Problemen kommt es aktuell bei der faktischen Rechtsdurchsetzung vor allem im Zusammenhang mit Beleidigungen und anderen Eingriffen in die Persönlichkeitsrechte aufgrund der Extraterritorialität des Internet und der Kosten der Rechtsverfolgung. Zudem bestehen Rechtsunsicherheiten und Rechtsschutzdefizite für Internet Service Provider bei Maßnahmen wie Netzsperren zur Durchsetzung des Urheberschutzes ohne richterliche Kontrolle.

Die Bedeutung der Plattformbetreiber für die gesellschaftliche Kommunikation hat auch in Österreich zu anhaltenden Diskussionen über eine Verschärfung der Haftung oder zur Erwei-

terung der Verpflichtung zur Löschung rechtswidriger oder anstößiger Inhalte geführt. Zuletzt bekam das Thema „Hass im Netz“ verstärkte öffentliche Aufmerksamkeit aufgrund sexistischer, frauenverachtender und abwertender Kommentare. Die österreichische Bundesregierung plant ein „digitales Vermummungsverbot“ das im Frühjahr 2019 mit einem „Gesetz für Sorgfalt und Verantwortung im Netz“ zur Begutachtung vorgelegt wurde. In Bezug auf Fragen zu Providerhaftung und Löschverpflichtung erwartet die österreichische Politik sich primär eine Lösung auf europäischer Ebene.

Die Europäische Union entwickelt ihre Politik zum Thema Internet-Plattformen u.a. in den Politikfeldern digitaler Binnenmarkt, Wettbewerb, audiovisuelle Medien, Urheberschutz sowie Kinder- und Jugendschutz. Dabei betont die EU-Kommission einerseits die Selbstverantwortung der Plattformen, setzt für etliche Problembereiche auf selbstregulatorische Lösungen und entwickelt selbst die entsprechenden Initiativen, z.B. zur Bekämpfung von Hassrede und Desinformation. Gleichzeitig übt die Kommission zunehmenden regulatorischen Druck auf die Industrie aus und entwickelt rechtliche Vorgaben (Jugendschutz, Urheberschutz, Terrorismusbekämpfung), die direkt oder mittels Ko-Regulierung umzusetzen sind.

Die Reformstrategien der EU werden kontrovers diskutiert und legislative Interventionen, wie die Urheberrechtsreform, werden aufgrund möglicher Beschränkungen der freien Meinungsäußerung kritisiert. Problematisiert wird zudem der Druck in Richtung Selbstregulierung, weil eine „freiwillige Inhaltsregulierung“ durch Unternehmen anders als staatliche Regulierung sich der grundrechtlichen Beurteilung durch Gerichte entzieht.

Aus der Analyse des Status quo der rechtlichen Rahmenbedingungen und der Praxis der Inhaltsregulierung auf Internet-Plattformen lassen sich zwei übergreifende Trends erkennen:

Zum einen führt der aktuelle Regulierungsrahmen zu einem *doppelten und damit sehr starken Privileg für Plattformbetreiber* (vgl. Gillespie 2018) – mit weitreichenden Implikationen: Bislang liegen die Kompetenzen für die Inhaltsregulierung fast ausschließlich (mit Ausnahme gerichtlicher Anordnungen zur Löschung illegaler Inhalte) bei den Plattformen selbst. Sie genießen dafür als erstes großes Privileg eine Haftungsregelung, die sie nicht dazu verpflichtet, Inhalte auf ihren Plattformen aktiv zu kontrollieren. Praktisch behindern geltende Regelungen sogar die Inhaltskontrolle, weil Haftungsprivilegien verwirkt werden, wenn Plattformen durch Inhaltskontrolle Kenntnis über Inhalte erlangen. Die Plattformen genießen darüber hinaus als zweites großen Privileg die Freiheit und Erlaubnis, Inhalte auf ihren Plattformen ganz nach eigenem Ermessen und eignen Standards zu löschen, zu filtern und zu kuratieren – und so ihre eigenen inhaltlichen Regeln weitgehend ohne Aufsicht und Rechenschaftspflicht zu setzen und durchsetzen zu können.[93]

[93] Vgl. auch Bunting (2018): „The ECD has created an industry in ‘notice and take down’ mechanisms as the main form of platform oversight, but leaves platforms’ much wider role in regulating online content largely to their discretion.“

Die Implikationen der doppelten Privilegierung für Plattformunternehmen sind weitreichend. Das Privileg, Inhalte nicht kontrollieren zu müssen, hat dazu beigetragen, dass im Internet vielfältige illegale und problematische Inhalte verfügbar sind, die sich negativ auf gesellschaftliche Entwicklungen auswirken. Das Privileg, Inhalte nach eigenem Ermessen moderieren zu dürfen, wiederum wirkt sich negativ auf die Meinungsfreiheit aus. Beide Privilegien führen in Kombination zu einer starken Asymmetrie zu den Anforderungen, denen andere Akteure im gesellschaftlichen Kommunikationssektor unterliegen, wie z.B. Zeitungsherausgeber und Fernsehveranstalter. Dies führt nicht nur zu ungleichen Wettbewerbsbedingungen, sondern auch zu ungleichen Schutzstandards für das Publikum, je nachdem über welche Plattform die Inhalte verbreitet werden.[94] Gillespie (2018, xiif.)[95] leitet aus der doppelten Privilegierung die Notwendigkeit für begleitende öffentliche Verpflichtungen hinsichtlich Fairness, Transparenz und Menschenwürde ab, die auf Grundsätze eines verantwortungsorientierten Governance-Arrangements verweisen.

"Safe harbor, including the right to moderate in good faith and the permission not to moderate at all, was an enormous gift to the young Internet industry. Over the history of U.S. regulation of the media and telecommunication industries, gifts of this enormity were always fitted with a matching obligation to serve the public. (…) The gift of safe harbor could finally be fitted with public obligations – not external standards for what to remove, but parameters for how moderation should be conducted fairly, publicly, and humanely." (Gillespie 2018, xiif.)

Zum anderen deuten aktuelle Reformen und Reformüberlegungen auf unterschiedlichen Ebenen des Regulierungssystems darauf hin, dass diese doppelte Privilegierung erodiert. Durch regulatorischen Druck, v.a. auf EU-Ebene, werden Internet-Plattformen zu Maßnahmen im Bereich der Inhaltsregulierung gedrängt und stärker in die Verantwortung genommen. Damit steigt jedoch gleichzeitig auch der Einfluss der Internet-Plattformen auf gesellschaftliche Kommunikation und es steigt die Bedeutung der Plattformen als politischer Machtfaktor (Helberger 2018).[96] Daher besteht ein Bedarf an einer stärkeren Verteilung von Verantwortlichkeit und der Implementierung geeigneter „checks and balances", um diese Macht zu kontrollieren.

[94] Waterson (2018), https://www.theguardian.com/media/2018/sep/17/ofcom-to-push-for-regulation-of-social-media-platforms

[95] Gillespie (2018) bezieht sich dabei auf die geltenden Regelungen zur Haftungsprivilegierung in den USA mit Section 230 des Communications Decency Act (CDA).

[96] Siehe z.B. pointiert formuliert hier: https://twitter.com/nhelberger/status/1064258139384627200

4. Grundsätze und institutionelle Komponenten eines verantwortungsorientierten Governance-Arrangements

Im vierten Arbeitsschritt erfolgt eine Identifikation von Grundsätzen und Kernelementen für die Ausgestaltung eines institutionellen Governance-Arrangements für Inhaltsregulierung auf Internet-Plattformen. Mit Hilfe von Literaturanalysen werden Vorschläge identifiziert, die in der Diskussion über Inhaltsmoderation gemacht werden. Zu solchen Vorschlägen für Prinzipien zählen z.B. Transparenz, Verhältnismäßigkeit, Berufungsoptionen und Vorschläge für adaptive Governance. Solche Vorschläge für Grundsätze werden gesammelt und dahingehend überprüft, welche institutionellen Kernelemente für ein Governance-Arrangement damit verbunden sind oder sich daraus ableiten lassen. Zu solchen möglichen institutionellen Kernelementen zählen etwa Verhaltenskodizes, Ombudsstellen und Beschwerdeverfahren, unabhängige Aufsichtsstellen zur externen Kontrolle sowie weitere Media-Accountability-Instrumente (Fengler, Eberwein & Leppik-Bork 2011; Eberwein, Fengler & Karmasin 2018), deren Potenzial im Bereich der Governance von Internet-Inhalten überprüft werden kann. Im Ergebnis liefert Abschnitt vier einen Überblick zu Prinzipien und institutionellen Eckpfeilern für ein verantwortungsorientiertes Governance-Arrangement für die Inhaltsmoderation auf Internet-Plattformen.

4.1 Rahmenkonzept & Leitprinzip: Verantwortlichkeit (Accountability)

Den Rahmen der folgenden Analyse bildet das Konzept der „Verantwortlichkeit" (Accountability), das gleichzeitig auch als zentraler Grundsatz in einem prinzipienorientieren Governance-Ansatz fungiert.

Umgangssprachlich wird Verantwortlichkeit häufig mit organisatorischen Kompetenzbereichen assoziiert, für die von bestimmten Personen Verantwortung übernommen wird. In einem engeren, rechtlichen Sinn wird Verantwortlichkeit vor allem durch Haftung (liability) und Schadenersatz hergestellt.[97] Dafür muss das Verantwortungssubjekt eine gewisse Kontrolle über die Sachverhalte haben – in dem Sinn, dass es einen Schaden verursacht oder mitverursacht oder sich in einer Position befindet, um einen Schaden abzuwenden. Verantwortlichkeit kann aber auch als Mechanismus verstanden werden (Bovens 2010), der über die organisatorische Festlegung von Zuständigkeits- und Verantwortungsbereichen und rechtliche Verursa-

[97] "Legally, accountability manifests itself through the concept of liability to provide a remedy (such as damages). The law usually imposes liability on the person who is in a position to prevent harm or mitigate a risk (for example through insurance)" (Council of Europe Committee of experts on internet intermediaries 2018).

chungs- und Haftungsfragen hinausgeht und eine starke kommunikative Komponente beinhaltet. Dabei beschreibt der Begriff Accountability die Bereitschaft oder Verpflichtung, Verantwortung zu übernehmen, verantwortungsvoll zu handeln und Rechenschaft abzulegen. Bovens (2007, 450) definiert Accountability als „a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor has to face consequences."

In der politikwissenschaftlich orientierten Accountability-Forschung wird in der Regel gefragt, wer wem gegenüber wofür verantwortlich ist (Scott 2000, 41).[98] Diese Frage stellt sich auch in Hinblick auf die Inhaltsregulierung durch die Betreiber von Internet-Plattformen und mitunter durch staatliche Inhaltsregulierung. Welche Grundsätze und welche institutionellen Kernkomponenten soll ein Governance-Arrangement umfassen, das die Verantwortlichkeit von Internet-Plattformen gegenüber Nutzern, Öffentlichkeit und Politik in den Mittelpunkt rückt? In funktionaler Hinsicht kann Accountability verschiedene Zwecke erfüllen. So ist mit Fisher (2014, 510) Accountability ein Mittel, um das Verhältnis zwischen Institutionen zu konkretisieren, Funktionen und Zuständigkeiten abzugrenzen, Macht zu kontrollieren, Legitimität zu stärken und Demokratie zu fördern. Ein verantwortungsorientierter Governance-Ansatz kann zur Erreichung solcher Ziele beitragen.

"The key challenge is to foster accountability while preserving the right to freedom of expression – even when the views are challenging, radical, and are offensive to some – in the face of pressure to protect the public from disinformation and the promotion of hatred or harm." (LSE 2018, 33)

4.1.1 Grundsätze für die Inhaltsregulierung

Meinungsfreiheit

Das Recht auf freie Meinungsäußerung, die Freiheit der Meinung und die Freiheit zum Empfang und zur Mitteilung von Nachrichten oder Ideen muss als Leitprinzip für eine liberale Gesellschaft erhalten bleiben (vgl. LSE 2018, 35; Ofcom 2018, 27). Inhaltsregulierung muss sich an diesem Prinzip orientieren. „Staatliche" Eingriffe durch öffentliche Behörden müssen zwingend gesetzlich vorgesehen sein, einem legitimen Zweck dienen, einem bestimmten öffentlichen Interesse oder den Rechten anderer, und sie müssen verhältnismäßig sein. Standards und Prozesse der Inhaltsregulierung durch die „privaten" Plattformen liegen zwar in deren eigenem Ermessen, sollten aber ebenfalls den Grundsatz der Meinungsfreiheit beachten.

---

[98] Als Relationselemente von Verantwortung gelten mit Sombetzki (2016, 3f.) die Verantwortungsträgerinnen (Subjekte), Handlungen und Handlungsfolgen, für die Verantwortung übernommen wird (Objekte), Verantwortungsinstanzen wie Gerichte mit mehr oder weniger starken Sanktionsfähigkeiten, Adressaten als Betroffene der fraglichen Verantwortlichkeit (z.B. Geschädigte) und jene normativen Kriterien, die den Maßstab für die Zuschreibung von Verantwortung bilden, z.B. in Form von Werten, Prinzipien, Geboten, Gesetzen oder Pflichten.

Primat der Plattformen bei Inhaltsregulierung

Die operative Tätigkeit der Inhaltsregulierung auf Plattformen durch löschen, sperren, moderieren, kuratieren muss weiterhin zentral in der Kompetenz und Verantwortung der Plattformen selbst liegen. Zur Sicherung von Meinungsfreiheit und Vertrauen muss die Inhaltsregulierung vor direkten staatlichen Eingriffen und anderen Formen der Manipulation durch Dritte geschützt werden. Mögliche Re-Regulierungen dürfen kein Einfallstor für staatliche Zensur eröffnen. In vor- und nachgelagerten Bereichen der Inhaltsregulierung (Meldungen, Berufung, Aufsicht, Kennzeichnung) kann eine stärkere Einbindung gesellschaftlicher Akteure zu einer besseren Verteilung von Verantwortlichkeit genutzt werden.

Legitimität und Verhältnismäßigkeit

Inhaltsregulierung durch Filterung und Kuratierung auf Internet-Plattformen kann das Recht auf Meinungsäußerung beschränken, als Zensur wirken, die Wahrnehmung der Realität verändern und das Vertrauen der Menschen unterminieren. Deshalb muss eine Inhaltsregulierung auf Internet-Plattformen den Prinzipien der Legitimität und Verhältnismäßigkeit folgen. Für die Sicherung der Legitimität muss Klarheit über die Ziele der Inhaltsregulierung hergestellt werden, denen die Eingriffe dienen. Willkürliche Eingriffe müssen verhindert werden. Eingriffe im Rahmen der Inhaltsregulierung müssen verhältnismäßig sein. Plattformbetreiber haben vielfältige Instrumente zur Inhaltsregulierung entwickelt, die eine unterschiedliche Eingriffsintensität aufweisen (Löschung, Sperrung, De-Priorisierung, Kommentierung, Kennzeichnung, etc.). Die Intensität der Eingriffe muss den Gefahren angemessen sein. Das bedeutet praktisch beispielsweise Maßnahmen zu setzen, um „overblocking" zu verhindern.

Transparenz

Normen, Verfahren und Ergebnisse/Entscheidungen der Inhaltsregulierung müssen transparent und nachvollziehbar sein. Für Konsumenten muss klar sein, wie Inhaltsregulierung erfolgt und was sie von den Plattformbetreibern in Bezug auf Inhaltsregulierung erwarten können (Bunting 2018, 13). „Es kann nicht sein, dass (ausländische) private Unternehmen Kommunikationsvorschriften erlassen oder nach firmeneigenen Regeln agieren (Eingriffe, Löschungen etc.), ohne dass eine angemessene Öffentlichkeit darüber hergestellt wird" (Jarren 2018). Transparenz gilt daher als grundlegende Anforderung in einem verantwortungsorientierten Governance-Arrangement.

Transparenz über Normen und Entscheidungskriterien

Inhaltsregulierung durch Filtern und Löschen muss auf klaren Normen und Standards basieren, mit denen festgelegt ist, welche Inhalte nicht erlaubt oder akzeptiert sind. Dies gilt zum

einen für gesetzlich unerlaubte Inhalte, die durch die Gesetzgebung festgelegt sind.[99] Aber auch Plattformbetreiber müssen ihre Normen und Standards (Unternehmensregeln, Community-Leitlinien) fest- und offenlegen. Plattformen sollen transparent und detailliert darüber informieren, welche Inhalte nicht erlaubt sind, und sie sollen offenlegen, wie (nach welchen Kriterien) Entscheidungen über Zulässigkeit oder eine Sperre der Inhalte getroffen werden. Auch Inhaltsregulierung durch Kuratierung und Targeting in der Werbung muss transparent sein. Plattformen müssen die Nutzer darüber informieren, nach welchen Kriterien ihnen Inhalte und Werbung angezeigt werden. Die Gesellschaft muss über die Funktionsweise der Kuratierung informiert werden.[100]

Transparenz und Information gegenüber Nutzern bei Einzelfallentscheidungen („notice")

Wenn Maßnahmen gegen Inhalte ergriffen werden, dann sollen Nutzer darüber informiert werden. In den Santa Clara Principles (2018) wird vorgeschlagen, Nutzer darüber in Kenntnis zu setzen, welche ihrer Beiträge gesperrt wurden, welche konkreten Normen (gemäß Community-Standards, Nutzungsrichtlinien) mit den Beiträgen verletzt wurden, welche Gründe zu einer Maßnahme geführt haben, wie der Inhalt als Verletzung der Norm identifiziert wurde (z.B. manuelles oder automatisiertes Flagging, Anordnung durch Behörden) sowie eine Erläuterung zum Berufungsverfahren darlegen. Auch jene Nutzer, die problematische Inhalte melden, sollen Informationen über die Ergebnisse der Moderationsverfahren erhalten.

Faire Verfahren und Beeinspruchungsmöglichkeiten („appeal")

Im Rahmen der Inhaltsmoderation bestehen vielfach Interessengegensätze und Grauzonen hinsichtlich zulässiger und nicht zulässiger Inhalte. Plattformen müssen deshalb Möglichkeiten schaffen, getroffene Entscheidungen zu beeinspruchen, insbesondere, wenn es sich bei Entscheidungen um gesperrte Konten oder gelöschte Beiträge handelt. Mindeststandards für Berufungsverfahren umfassen gemäß der Santa Clara Principles (2018) u.a. eine Begutachtung der inkriminierten Beiträge durch unabhängige Personen oder Panels, Möglichkeiten zusätzliche Argumente ins Treffen zu führen, die im Verfahren berücksichtigt werden, und das Recht auf eine verständlich formulierte und begründete Entscheidung über das Ergebnis des Einspruchsverfahrens. Von verschiedenen Stakeholdern wird zudem gefordert, dass solche Einspruchsverfahren außerhalb der Organisationsstrukturen der Plattformen durch *externe, unabhängige Institutionen* abgewickelt werden.[101]

---

[99] Vgl. Ofcom (2018): „Our experience indicates that protection and assurance can be effectively achieved where Parliament sets clear statutory objectives as to the kind of harms to be prevented. These objectives reflect societal norms and set clear standards covering the types of content and conduct that are likely to cause harm. Regulated parties are required to adopt practices or procedures designed to secure these objectives."

[100] Siehe z.B. Statement von Ofcom Chief Executive Sharon White: „Platforms should also be asked to improve transparency so audiences would know why they were targeted with certain material" (https://www.ft.com/content/a16935a4-bb39-11e8-94b2-17176fbf93f5).

[101] „In the long term, independent external review processes may be an important component for users to be able to seek redress." (Santa Clara Principles)

Transparenz durch Offenlegung von Moderationsdaten („evidence & numbers")

Grundlage für Transparenz ist der Zugang zu Daten über die Praxis der Inhaltsmoderation. Als Grundprinzip sollte gelten, dass Plattformen Daten zur Inhaltsmoderation offenlegen. In den Santa Clara Principles wird vorgeschlagen, die Anzahl gesperrter Beiträge und Konten nach folgenden Informationen zu veröffentlichen: Formate (Text, Video, etc.), Kategorien von Verstößen (Gewalt, Pornographie, etc.), Beschwerdequellen (Nutzer, Gerichte, Trusted Flagger, etc.), sowie der Herkunft/Standort der gesperrten Konten und der Beschwerdeführer. Zur Verbesserung der Transparenz sollten solche Daten über Inhaltsregulierung regelmäßig von den Plattformbetreibern zur Verfügung gestellt werden. Die Offenlegung der entsprechenden Daten ist eine der Formen, um Rechenschaft abzulegen, und eine Voraussetzung für die Überprüfung der Inhaltsregulierung (Kontrolle, Audits).

Unabhängige Kontrolle, Überprüfung und Evaluierung („audit, review")

Normen, Prozesse/Verfahren und Ergebnisse der Inhaltsregulierung sollen Gegenstand von regelmäßigen Überprüfungen sein. Mit regelmäßigen Überprüfungen soll festgestellt werden, ob Normen adäquat ausgewählt und definiert sind, ob Verfahren demokratischen Standards entsprechen und ob mit den Ergebnissen der Inhaltsregulierung auch die angestrebten Ziele erreicht werden. Die Überprüfungen und Evaluierungen sollen von Stellen durchgeführt werden, die von den Plattformen unabhängig sind. Auch andere einseitige Einflüsse sollen vermieden werden. Mögliche Aufsichts- und Regulierungsstellen müssen auch strukturell unabhängig von der Regierung sein, inklusive ihrer Personalauswahl und Finanzierung (LSE 2018, 35). Für Audit und Evaluierung schlägt Gillespie (2018, xv) vor, Systeme zu implementieren, die Zugang zu Moderationsdaten und regelmäßige Audits durch Forscherinnen und Forscher, Journalistinnen und Journalisten und sogar Nutzer ermöglichen.

Stakeholder-Einbindung und Interessenabwägung

Selbstregulierung der Industrie steht häufig im Verdacht, nur die Industrieinteressen zu bedienen und öffentliche Interessen zu vernachlässigen. In der Inhaltsregulierung sollte jedoch ein Ansatz verfolgt werden, der „fair" ist und auch öffentliche Interessen sowie legitime Interessen und Rechte Dritter beachtet (Bunting 2018, 14). Eine Möglichkeit, für eine Kontrolle und eine faire Abwägung von Interessen zu sorgen, ist die Einbindung von Stakeholdern in Regulierungsarrangements. Zu den Stakeholdern können, je nach Problemstellung, staatliche Institutionen, Konsumentenverbände, zivilgesellschaftliche Organisationen oder andere Anspruchsgruppen zählen. Die Einbindung kann strukturell in Aufsichtsgremien erfolgen – oder punktuell, zum Beispiel bei der Erstellung von Verhaltenskodizes. Für die Inhaltsregulierung auf Internet-Plattformen wird die Einbindung von zivilgesellschaftlichen Organisationen in die Reformprozesse gefordert, mit denen die Selbstregulierung oder Ko-Regulierung etabliert und weiterentwickelt wird (LSE 2018, 35). Zudem wird gefordert, unterschiedliche An-

spruchsgruppen in die Träger- und Entscheidungsstrukturen von Kontrolleinrichtungen einzubinden.[102] Dabei kann es auch notwendig sein, zivilgesellschaftliche Organisationen mit den Mitteln für die Mitwirkung an strategischen und operativen Tätigkeiten auszustatten.

Demokratische Ansätze & Nutzereinbindung

Aktuell liegt die Hauptverantwortung für Inhaltsmoderation vor allem bei den Plattformen selbst, die dabei jedoch von Nutzern (Meldungen, Flagging), menschlichen Moderatoren und zivilgesellschaftlichen Organisationen unterstützt werden. Gillespie (2018, viiif.) entwickelt Vorschläge, diese Ressourcen auszubauen und besser zu nutzen, um damit eine stärkere Verteilung von Einfluss und Verantwortung zu ermöglichen. So soll das Engagement der Nutzer verwendet werden, um die kollektive Kennzeichnung von Inhalten zu ermöglichen (collective lenses; crowdsourced tags; shared blocklists[103]). Dabei könnte vor allem die Art der Inhalte, aber auch die Einschätzung von Gefährdungen beurteilt werden. Darauf aufbauend könnten aggregierte Daten produziert werden, die Nutzern dabei helfen, ihren eigenen Konsum zu steuern („Platforms should let flagging accumulate into actionable data for users", ebd., ix). Die Kategorien und Zuordnungen könnten in moderierten Foren offen für Diskussion und Weiterentwicklung sein. Kollektive Kennzeichnung könnte auch dabei helfen illegale Inhalte und problematisches Nutzerverhalten zu identifizieren. In Summe würde das vielfältige Nutzerengagement genutzt und Moderationsverfahren weniger zentral und stärker kollektiv und demokratisch organisiert.

4.1.2 Grundsätze für das Governance-Arrangement

Subsidiarität und Kontextsensibilität

Internet-Plattformen agieren global, aber ihre Entscheidungen und Programme werden regional wirksam. Es ist wichtig eine stärkere Kontextsensibilität der Inhaltsregulierung zu erreichen und Arrangements zu fördern, die auf nationale Besonderheiten Rücksicht nehmen, ohne dabei die Funktionsfähigkeit der Plattformen zu gefährden. Deshalb sollten Entscheidungen über Inhaltsstandards auch von jenen mitentschieden werden, die sie betreffen. „Das wird oft bedeuten auf nationaler Ebene" (LSE 2018, 35). Zu untersuchen ist, welche institutionellen Arrangements so eine nationale Rückkoppelung der Inhaltsregulierung ermöglichen.

Adaptive Governance und Überprüfung der Reformprozesse („adaptability")

Die Reform des institutionellen Arrangements der Inhaltsregulierung im Internet und auf Internet-Plattformen wird ein iterativer Prozess sein. Es wird nicht möglich sein, ein komplett neues, finales Arrangement zu entwickeln und zu implementieren, u.a. weil sich Nutzungsverhalten, Erwartungen und Dienste-Angebote wandeln. Laufende Anpassungen werden

[102] Siehe z.B. die Forderung nach Unabhängigkeit und Stakeholder-Einbindung in einem Social Media Council auf Basis einer Selbstregulierung (Article 19, 2018, 21) oder die Forderung nach einer unabhängigen Regulierungsbehörde, Independent Plattform Agency (LSE 2018).

[103] Gillespie (2018, ix) unter Verweis auf Geiger (2016).

notwendig sein (Ofcom 2018, 4), und ein Reformprozess, auch mit Versuchen und Fehlschlägen (trial and error), muss möglich sein. Deshalb muss der Reformprozess laufend überprüft werden, um herauszufinden welche institutionellen Governance-Komponenten und welche Maßnahmen sich bewähren und wo es weiteren Reformbedarf gibt. Ansätze wie „adaptive Governance“ und „antizipatorische Governance“ (Dzakula 2018) sind durch Strukturen gekennzeichnet, die regulatorische Flexibilität für Innovationsprozesse unter einer Aufsicht bieten, die auf einem kontinuierlichen Dialog zwischen Regulatoren und Industrie und Anpassungen der Regulierung basiert. Dabei sollen technologische Veränderungen ebenso berücksichtigt werden wie neue Herausforderungen, die Funktionsweise der Internet-Plattformen, die Praxis der Inhaltskontrolle und die sich immer wieder ändernden Formen von Belästigung, Hass, Desinformation und anderen Gefahren (Gillespie 2018, xv). Überprüft werden sollen auch die rechtlichen Rahmenbedingungen der Inhaltsregulierung wie die geltende Haftungsprivilegierung gemäß der E-Commerce-Richtlinie.

Differenzierte Governance-Ansätze

Dienste im Internet haben unterschiedliche Reichweiten, unterschiedliche Formate und Geschäftsmodelle und sie bergen unterschiedlich hohe Gefahren und Risiken. Mitunter unterliegen ähnliche Angebote je nach Anbieter und Abspielkanal sehr unterschiedlichen Regeln (Ofcom 2018). Vielfach wird argumentiert, dass Größenunterschiede und Unterschiede in Geschäftsmodellen der Unternehmen in Regulierungsansätzen berücksichtigt werden sollen (z.B. Bunting 2018, 14). Hohe Anforderungen sollten vor allem für große Anbieter gelten, denn von ihnen gehen aufgrund der größeren Reichweite auch größere Gefahren aus. Für sie sollten höhere Standards gelten und sie sollten größere Anstrengungen unternehmen, um Risiken zu begegnen (Perrin & Woods 2018).

Steuerungsansätze

Im Rahmen der Inhaltsregulierung geht es u.a. darum, wie die regulatorischen Aufgaben zwischen staatlichen und privaten Akteuren verteilt werden und welche Anforderungen und Vorgaben von staatlicher Seite für die privaten Plattformbetreiber gemacht werden. Dabei kann zwischen unterschiedlichen Arten von Vorgaben differenziert werden, die auch mit unterschiedlichen Steuerungsansätzen einhergehen:

- *Regel-zentrierte Ansätze* (command-and-control) basieren auf konkreten substanziellen inhaltlichen Vorschriften mittels Verboten oder Geboten (z.B. keine Nacktheit) und/oder konkreten Vorgaben, mit welchen Maßnahmen und Mitteln die Verpflichtungen umgesetzt werden sollen (z.B. Upload-Filter).
- *Ergebnis-orientierte Ansätze* (outcome-based) legen Ziele fest und überprüfen, ob und inwieweit die Ziele erreicht werden. Sie überlassen die Wahl der Mittel zur Zielerreichung den Marktakteuren.
- *Prozedurale Ansätze* (procedural standards) legen Ziele fest und formulieren grundsätzliche Anforderungen für den Regulierungsprozess, z.B. dass Standards veröffentlicht

werden und Berufungsmöglichkeiten vorgesehen sind. Durch den Fokus auf grundsätzliche Standards wird der Ansatz auch als „principles-based approach“ bezeichnet.

Die Ansätze haben jeweils spezifische Vor- und Nachteile. Regelbasierte Ansätze schaffen beispielsweise starke Rechtssicherheit. Ergebnisorientierte Ansätze bieten Flexibilität bei der Umsetzung. Plattformunternehmen schlagen z.B. vor, ihre Inhaltsregulierung anhand von Performance-Indikatoren zu überprüfen (Zuckerberg 2018).[104] Im Feld der Inhaltsregulierung im Internet wird jedoch auch argumentiert, dass prozedurale Ansätze angemessen sind (z.B. Bunting 2018). Konkrete inhaltliche Verpflichtungen oder Vorschriften können die Meinungsfreiheit beschränken und wie staatliche Zensur wirken. Vorgaben hinsichtlich der einzusetzenden Mittel können aufgrund des technologischen und wirtschaftlichen Wandels schnell obsolet sein. Deshalb sollen Unternehmen selbst entscheiden, mit welchen Maßnahmen sie die Ziele erreichen. Allerdings sind viele der inhaltsbezogenen Ziele nicht operationalisierbar und messbar und die Zielerreichung nur schwer zu überprüfen. Deshalb sind rein ergebnisorientierten Ansätzen Grenzen gesetzt, die für prozedurale Ansätze so nicht bestehen.

Gefordert werden beispielweise *Basisstandards* für die Inhaltsmoderation, ohne dass dabei gleich bestimmte Moderationspraktiken festgelegt werden. Als Basisstandards bezeichnet werden u.a. minimale Standards hinsichtlich besonders problematischer Inhalte, festgelegte Zeitrahmen für Moderationsmaßnahmen oder obligatorische Mechanismen für den Umgang mit Einsprüchen (Gillespie 2018b, 214). Perrin und Woods (2018) schlagen demgegenüber einen Regulierungsansatz vor, der auf einer positiven Sorgfaltspflicht („positive statutory duty of care“) von Plattformbetreibern gegenüber den Nutzern ihrer Dienste basiert und v.a. grundlegende Standards zur Reduktion von Gefahren festlegt.[105] Dabei werden Plattformen mit öffentlich zugänglichen Räumen wie Bars oder Erlebnisparks verglichen, deren Bertreiber für die Sicherheit der Besucher sorgen sollen und angemessene Maßnahmen setzen müssen, um Gefahren zu verhindern. Laut dem House of Lords (2019, 55), das diesen Vorschlag unterstützt, soll die Einführung der positiven Sorgfaltspflicht („duty of care“) durch die Plattformen dazu beitragen eine Kultur des Risikomanagements auf allen Stufen der Entwicklung und des Betriebs der Plattformdienste zu erwirken.

4.2 Institutionelle Kernkomponenten (Optionen, Ideensammlung)

Basierend auf den oben identifizierten Grundsätzen werden nun in einem weiteren Schritt institutionelle Komponenten für die Ausgestaltung eines verantwortungsorientierten Gover-

[104] „I believe the ideal long term regulatory framework would focus on managing the prevalence of harmful content through proactive enforcement. This would mean defining the acceptable rates of different content types. (...) A good starting point would be to require internet companies to report the prevalence of harmful content on their services and then work to reduce that prevalence. Once all major services are reporting these metrics, we'll have a better sense as a society of what thresholds we should all work towards“ (Zuckerberg 2018).

[105] Perrin und Woods schlagen vor, die Gefahren auf Plattformen konkret zu definieren und zu messen, um daraus adäquate Maßnahmen abzuleiten, mit denen Gefahren reduziert werden sollen (Perrin & Woods 2018).

nance-Arrangements für die Moderation von Internet-Inhalten identifiziert und kurz beschrieben. Diese Kernkomponenten werden aus den Grundsätzen abgeleitet und aus der einschlägigen Literatur identifiziert. Ziel dabei ist es, unterschiedliche Vorschläge zu sammeln, und ihre Vor- und Nachteile, ihre Potenziale und Grenzen darzulegen.

Dabei ist es wichtig zu betonen, dass die Analyse auf den Bereich der Inhaltsmoderation auf Internet-Plattformen beschränkt ist. Andere wichtige Regulierungsfelder wie etwa die Werbe- und Geschäftsmodelle, der Schutz der Privatsphäre und das Problem der Marktmacht der Plattformen werden mit der Analyse nicht umfasst. Die im Folgenden diskutierten institutionellen Vorschläge für Inhaltsmoderation sind deshalb kein Ersatz für notwendige Regulierung und Rechtsaufsicht in anderen Problemfeldern wie Datenschutz und Wettbewerb.

Zudem ist es wichtig zu betonen, dass es sich bei den Vorschlägen um „institutionelle" Optionen im Bereich der Inhaltsmoderation handelt, die „national" umgesetzt werden können. In der Analyse nicht berücksichtigt sind deshalb z.B. Möglichkeiten für Inhaltsregulierung die ausschließlich bei den Plattformen selbst, bei den Technologien oder beim Nutzerverhalten ansetzen. Dazu zählt beispielsweise die wichtige Option der Gegenrede (Counter-Speech) durch andere Nutzer bei Hasskommentaren, die zur Solidarisierung mit Opfern beitragen können und zeigen, dass Hass nicht die dominierende Meinung auf einer Plattform ist.[106] Anstatt Inhalte zu löschen kann so die Debatte verbreitert werden. In der Analyse nicht berücksichtigt sind auch plattformseitige technische Entwicklungen, mit denen z.B. die Nutzerkontrolle über Newsfeeds verbessert oder die Transparenz von Werbung erhöht werden soll. Zuletzt wurde von Plattformen u.a. in die Entwicklung solcher internationaler, technischer Instrumente investiert. Im Folgenden stehen jedoch Governance-Optionen im Zentrum die eine nationale Rückbindung der Inhaltsmoderation ermöglichen.

Zu den identifizierten Kernelementen zählen institutionelle Vorkehrungen für verantwortungsorientierte Kommunikation (4.2.1), Institutionen für Melde- und Beschwerdemanagement (4.2.2) sowie Institutionen für Aufsicht und Kontrolle (4.2.3).

4.2.1 Institutionen für verantwortungsorientierte Kommunikation

Accountability kann als Mechanismus verstanden werden, der auf Rede-und-Antwort-Stehen abzielt (Bovens 2007, 450) und damit eine starke kommunikative Komponente beinhaltet.[107] Internet-Plattformen ermöglichen zwar vielfältige Kommunikation und verlautbaren selbst ihre Pläne und Fortschritte, z.B. in Firmenblogs oder auf Entwicklerkonferenzen. Sie entziehen sich jedoch vielfach einer kritischen öffentlichen Diskussion. Die aktuelle Tour einiger CEOs durch diverse parlamentarische Befragungsausschüsse in den USA und Europa sind eher die

[106] Siehe auch den Saferinternet.at (2017): Aktiv gegen Hasspostings (Leitfaden), https://www.saferinternet.at/fileadmin/categorized/Materialien/Aktiv_gegen_Hasspostings_Leitfaden.pdf

[107] Vgl. Bovens (2007, 450).

Ausnahmen als die Regel. Plattformen sollten ihr Handeln generell stärker erklären und rechtfertigen und sich im Sinne einer kommunikativen unternehmerischen Verantwortlichkeit (Weder & Karmasin 2011; Karmasin & Litschka 2017) einer kritischen Öffentlichkeit und Diskussion stellen.[108]

Speziell weil die Inhaltsmoderation der Plattformen Auswirkungen auf nationale Öffentlichkeiten hat, müssen die Plattformen sich auch diesen nationalen Öffentlichkeiten gegenüber „verantworten", also Rede und Antwort stehen und Rechenschaft ablegen. Es muss klar sein, dass Plattformen nicht nur ihren einzelnen Nutzern gegenüber verantwortlich sind, sondern auch gegenüber den Öffentlichkeiten und Gesellschaften, die sie mit ihren Angeboten und ihrer Kuratierung prägen. Hürden bestehen, weil in Plattformunternehmen meist nur eine beschränkte Auswahl an Mitarbeitern offiziell Stellung nehmen darf und weil in der Regel auch Vorgaben hinsichtlich der Auswahl an Themen bestehen, die sich mit aktuellen Herausforderungen der Plattformkommunikation auf nationaler Ebene nicht notwendigerweise decken.

Nationale Plattformsprecher (Auskunftsverantwortliche)

Vorgeschlagen wird deshalb die Installierung von auskunftsberechtigen Personen, die auf Kritik reagieren und Rede und Antwort stehen. „Platforms should have public facing people who can answer to regulatory and other questions. Platforms could each be required to have a public ombudsman who responds to public concerns and translates those concerns to policy managers internally" (Gillespie 2018b, 214). Solche Ansprechpersonen sollten auch auf nationaler Ebene benannt werden, sich in nationale Debatten einbringen und sich kritischen Fragen stellen, und den Plattformen so im nationalen Kontext ein Gesicht geben. Dabei sollten diese Personen kommunikative Aufgaben erfüllen, die über die haftungstechnisch relevante Funktion eines inländischen Zustellungsbevollmächtigten oder Empfangsberechtigten für behördliche Auskunftsersuchen hinausgehen. So kann kommunikative Verantwortung gegenüber nationalen Öffentlichkeiten besser wahrgenommen werden. Eine Stärkung der kommunikativen Verantwortlichkeit im nationalen Kontext erfordert, dass Sprecher eingesetzt werden, die *bevollmächtigt* sind im Namen der Plattformen Auskunft zu geben und Fragen zu beantworten.

4.2.2 Institutionen für Melde- und Beschwerdemanagement

Des Weiteren besteht das Prinzip der Verantwortlichkeit auch gegenüber den Nutzern von Plattformen, die Inhalte erstellen, konsumieren und teilen. Verantwortlichkeit besteht dabei beim Umgang mit Inhalten durch die Plattformbetreiber sowie im Umgang mit Problemen und Beschwerden. Zu den möglichen institutionellen Kernkomponenten zählen Meldestellen, Inhaltsmoderation und Einrichtungen für Beschwerdeverfahren, für die auch Optionen für eine nationale Rückkoppelung bestehen.

[108] Zur „Corporate communicative responsibility" als Ziel und Mittel unternehmerischer Verantwortungswahrnehmung siehe Weder & Karmasin (2011); Zur Analyse von CSR als ökonomisches, ethisches und kommunikatives Konzept siehe Karmasin & Litschka (2017).

Nationale Melde- und Beratungsstellen

Individuelle Nutzer können Beschwerden in Bezug auf Inhalte entweder als Anzeigen bei staatlichen Stellen einbringen oder sie können Beschwerden über Meldebuttons direkt an die Plattformen richten. Zudem können sie sich an spezifische Meldestellen wenden, die außerhalb der Plattformen agieren, die oft auf bestimmte Typen von problematischen Inhalten spezialisiert sind und die in der Lage sind, die Rechtmäßigkeit der Inhalte vor dem Hintergrund der Gesetzlage und Community-Regeln professionell zu beurteilen, Beweise zu sichern und formgerechte Beschwerden an Plattformen zu übermitteln. Manche dieser Meldestellen fungieren als „Trusted Flagger", deren Beschwerden im Rahmen der Moderation durch die Plattformen Priorität eingeräumt wird.

Dem Netzwerk an zivilgesellschaftlichen und amtlichen Meldestellen wird allgemein eine gute Funktionsweise attestiert. Nachdem diese Meldestellen national verankert sind, bieten sie auch *nützliche, niedrigschwellige erste Anlaufstellen für Nutzerinnen und Nutzer*, die den Rechtsweg nicht beschreiten können oder wollen, und deren individuelle Beschwerden von den Plattformen nicht oder unbefriedigend behandelt werden. Anlauf-, Vorprüf- und Beratungsstellen verbessern die Möglichkeiten für Nutzer und Betroffene im Sinne des gewünschten User-Empowerment. Meldestellen mit „Trusted Flagger"-Status erfüllen daher sehr wichtige Funktionen als Anlauf- und (Vor-)Prüfstellen, und ihr Angebot könnte ausgebaut und weiter gestärkt werden, damit auch ressourcenintensive Beratungsfunktionen für Nutzer erfüllt werden können.

Moderationsteams mit regionalem/nationalem Kontextwissen

Vielfach wird darauf verwiesen, dass Inhaltsmoderation und Löschentscheidungen mit zu wenig Sensibilität für die jeweiligen Kontexte getroffen werden. Dies gilt insbesondere für automatisierte Löschverfahren und für Entscheidungen, die von Moderatorinnen und Moderatoren getroffen werden, die zentral organisiert und mit inhalts- und regionenspezifischen Kontexten nicht vertraut sind. Empfohlen wird daher, beispielsweise mehr Moderatorinnen und Moderatoren zu beschäftigen, die in ihrer Zusammensetzung diverser, besser ausgebildet und auch arbeitsrechtlich besser vor Ausbeutung geschützt sind als die Moderatorinnen und Moderatoren in Billiglohnländern wie Indien und den Philippinen (Gillespie 2018a, i).

Mit einer stärkeren Regionalisierung der Moderationsteams und der Moderation können regionale Kontexte besser berücksichtigt und Moderationsentscheidungen verbessert werden. Plattformen müssten dafür Moderatorinnen und Moderatoren mit nationalem/regionalem Kontextwissen beschäftigen (Ash et al. 2019, 11f). In Deutschland wurde unter dem Druck der Einführung des Netzwerkdurchsetzungsgesetzes von Facebook ein nationales Moderationsteam eingerichtet.[109] Zudem kann auf nationaler Ebene leichter auf Ausbildungs-, Arbeits- und Betreuungsstandards und auf psychologische Begleitung geachtet werden, die für die sehr belastende Moderationsarbeit wichtig sind.

[109] ORF.at (2018), https://oe1.orf.at/artikel/653750

Geteilte Verantwortlichkeit durch unabhängige Berufungsverfahren

Die Kompetenzen und Verantwortung für Inhaltsmoderation liegen derzeit, sofern es sich nicht um Löschaufträge staatlicher Stellen handelt, fast vollumfänglich ausschließlich in der Hand der Plattformen. Egal ob Inhalte gelöscht werden oder sichtbar bleiben – für Beschwerdeführer, deren Reklamationen zurückgewiesen werden, ist der Verweis der Plattformbetreiber auf „Community-Standards" meistens wenig befriedigend. Auch Beschwerden oder Einsprüche gegen Moderationsentscheidungen werden bislang – wenn überhaupt – durch andere Moderatorinnen und somit durch Akteure *innerhalb* der Plattformunternehmen abgewickelt. Damit liegt viel Macht und Verantwortung in den Händen weniger, und demokratische sowie rechtsstaatliche Prinzipien werden nicht beachtet.

Zuletzt hat Mark Zuckerberg im Fall von Facebook selbst kritisch angemerkt, dass Facebook nicht so viele wichtige Entscheidungen mit Auswirkungen auf die Redefreiheit selbst treffen soll und die Einrichtung einer unabhängigen Instanz für Berufungsverfahren angekündigt.

„As I've thought about these content issues, I've increasingly come to believe that Facebook should not make so many important decisions about free expression and safety on our own. In the next year, we're planning to create a new way for people to appeal content decisions to an independent body, whose decisions would be transparent and binding" (Zuckerberg 2018).

Tatsächlich wurde von Facebook (2019a; 2019b) ein Vorschlag für eine Berufungsinstanz vorgelegt. Ein externes „Oversight Board for Content Decisions" soll sich aus 40 internationalen, unabhängigen Expertinnen und Experten zusammensetzen und besonders umstrittene und kritische Moderationsentscheidungen überprüfen und bei Bedarf korrigieren.

Kritisiert wird u.a., dass ein solches Gremium letztlich wieder direkt an Facebook angebunden wäre und so keine wirklich unabhängige Entscheidung ermöglicht (Spielkamp 2018). Zudem stellt der Vorschlag auf ein zentral institutionalisiertes Gremium ab, bei dem die Gefahr besteht, dass das Verständnis für nationale Kontexte bei Moderationsentscheidungen fehlt. Das Board soll deshalb laut den Plänen von Facebook geographisch und kulturell ausgewogen besetzt werden und bei Bedarf nationale, sprachliche, kulturelle und sozio-politische Expertise einholen können (Facebook 2019, 4).

Ein alternativer oder komplementärer Ansatz wäre es, Berufungsverfahren von den Plattformen an national verankerte Gremien wie z.B. nationale Ombuds-/Schlichtungsstellen weiterzuleiten und zu delegieren. Eine solche Delegation bietet sich vor allem an, wenn von Entscheidungen speziell nationale Angelegenheiten berührt und nationale Öffentlichkeiten be-

troffen sind, die umsichtige Abwägungen und Kontextverständnis für nationale Begebenheiten erfordern. Eine Delegation von Berufungsverfahren könnte dabei nicht pauschal für alle, aber für ausgewählte, schwer zu entscheidende Fälle erfolgen. Diese könnten im nationalen Kontext in eigens dafür eingerichteten Gremien „verhandelt" werden. Dadurch würden insgesamt gleichzeitig mehrere positive Effekte ermöglicht (vgl. Kasten 1)

Ausgewählte Vorteile von nationalen Ombudsstellen

- Durch die Einrichtung von Gremien, die von den Plattformen unabhängig agieren, kommt es zu einer stärkeren Verteilung von Verantwortlichkeit und zu einer Machtreduktion. Unabhängige Berufungsgremien sind ein Schritt in Richtung besserer „checks and balances" in einem System in dem bislang die Plattformen selbst gleichzeitig als Legislative, Exekutive, Judikative und Presse agieren (vgl. Klonick 2017).
- Mit der Einrichtung von Berufungsverfahren bei Ombudsstellen werden rechtsstaatliche Prinzipien gestärkt.
- Für Plattformen wird die Verantwortlichkeit für Inhaltsregulierung und damit potenziell auch Haftung reduziert, wenn zusätzliche Instanzen die Legitimität von Inhalten beurteilen. Speziell eine Reduktion der rechtlichen Haftung könnte einen Anreiz für Plattformen bieten Ombudsstellen zu akzeptieren und mit Ihnen zu kooperieren. Plattformen hätten den Vorteil, dass sie die Rolle als „Richter über Meinungsfreiheit" teilweise abgeben könnten.
- Mit der Einrichtung von Gremien auf nationaler Ebene wird sichergestellt, dass in den jeweiligen Verfahren die nationalen Gegebenheiten (Kontextbedingungen) berücksichtigt sind.
- Mit der Einrichtung von unabhängigen Berufungsgremien auf nationaler Ebene könnten gleichzeitig sowohl Plattformen als auch Gerichte entlastet werden. Angesichts der steigenden Anzahl an Nutzern und strittigen Inhalten könnten sich nationale Ombuds-/Schlichtungsstellen als hilfreich erweisen.
- Bei einer breiten Zusammensetzung des Entscheidungsgremiums, die Perspektivenvielfalt ermöglicht, können Entscheidungen umsichtig und unter Abwägung der verschiedenen involvierten Interessen getroffen werden.
- Ombudsstellen können eine eigene Spruchpraxis entwickeln, die den Besonderheiten der Internet-Kommunikation Rechnung trägt (Gostomzyk 2019), und in der rechtliche wie auch ethische Normen Berücksichtigung finden.
- Mitunter könnten ausgewählte Fälle auch für öffentliche, gesellschaftliche Diskurse sorgen, in denen über die Grenzen der Akzeptanz von problematischen Inhalten und den adäquaten Umgang damit diskutiert wird. Durch solche Diskurse entsteht ein Beitrag zur Verantwortungskultur, die sich nur durch Dialoge entwickeln kann (Jarren 2018).[110]

Berufungsgremien auf nationaler Ebene

Für die Delegation von Berufungsverfahren von den Plattformen auf die nationale Ebene ist die Einrichtung von Institutionen notwendig, die als nationale Ombuds-, Schlichtungs-, Clearingstellen fungieren (vgl. ähnlich Ladeur 2014; Ladeur 2017; Gostomzyk 2019; Jarvis 2019).

[110] Vgl. auch Jarvis (2019): „The process of negotiating legal norms, which in matters online is now occurring inside private corporations, must occur instead in public, in courts, and with due process."

Für die konkrete *institutionelle Ausgestaltung* solcher Einrichtungen stellen sich etliche Herausforderungen, v.a. hinsichtlich Zusammensetzung, Auftrag und der Verbindlichkeit der Ergebnisse für die Plattformen (Entscheidung, Empfehlung), dem Grad einer staatlichen Involvierung und ggf. auch notwendiger rechtlicher Grundlagen.

Im Vereinigten Königreich wird aus verschiedenen Blickwinkeln die Option einer Ombudsstelle für den Umgang mit problematischen Inhalten vorgeschlagen. Bei der Idee einer Ombudsstelle geht es letztlich v.a. darum, bei problematischen Inhalten in strittigen Fällen zwischen Nutzern und den Plattformen zu vermitteln. Perrin und Woods (2018) machen Vorschläge für eine Ombudsstelle zur Streitbeilegung für strittige Inhalte, die jedoch nicht illegal sind. Eine solche Schlichtungsstelle sollte Beteiligten dabei helfen, Streit außergerichtlich beizulegen und damit Überkriminalisierung zu vermeiden. „In relation to speech that is not obviously criminal, we are also considering whether an ombudsman function would help individuals better resolve disputes – perhaps helping to avoid over-criminalisation of ‚robust speech'" (Perrin & Woods, 2018). Das Department for Digital, Culture, Media and Sport (DCMS) beschäftigt sich hingegen mit Vorschlägen für eine Ombudsstelle für Beschwerden über illegale Inhalte wie Missbrauch und Gewaltandrohung. Die für Kinderangelegenheiten zuständige Regierungskommissarin schlägt eine Ombudsstelle für illegale Postings vor, an die sich speziell Kinder wenden können.[111]

In Deutschland sieht das Netzwerkdurchsetzungsgesetz (NetzDG) die Möglichkeit einer *regulierten Selbstregulierung* für Entscheidungen über rechtswidrige Inhalte vor. Dabei muss eine entsprechende Einrichtung vom Bundesministerium für Justiz anerkannt werden. Voraussetzung für eine Anerkennung sind Unabhängigkeit und Sachkunde der Prüfer, eine sachgerechte Ausstattung, eine interne Verfahrensordnung, die das Prüfverfahren regelt, eine Beschwerdestelle, an die sich Nutzer wenden können, deren Inhalte zu Unrecht entfernt wurden, Trägerschaft durch mehrere Netzwerke, sowie Offenheit für weitere Mitglieder.

In Australien wurden mit dem eSafety Commissioner[112] 2015 eine Behörde eingerichtet, der illegale Inhalte gemeldet werden können, die von sozialen Netzwerken nicht innerhalb von 48 Stunden gelöscht wurden. Der Commissioner kann Inhalte daraufhin prüfen und eine Löschung erwirken. Die Behörde wurde zunächst eingerichtet, um Kinder- und Jugendliche speziell vor Cyberbullying zu schützen. Ihr Mandat wurde aber 2017 zum Schutz Erwachsener erweitert. Kleinere Internet-Plattformen können sich an dem System beteiligen und werden aufgefordert Inhalte mit Cyberbullying zu entfernen, große Plattformen sind zur Teilnahme verpflichtet, und dem formalen notice and takedown Prozess unterworfen. Vergehen können mit Geldstrafen geahndet werden.[113]

---

[111] Bowcott & Gibbs (2017), https://www.theguardian.com/technology/2017/aug/22/uk-considers-internet-ombudsman-to-deal-with-abuse-complaints

[112] https://www.esafety.gov.au/

[113] Enhancing Online Safety Act 2015.

In Österreich wurde im Auftrag des Verbandes der Internet Service Provider (ISPA) ein Konzept einer „Clearingstelle“ zur Beurteilung von Netzsperren im Zusammenhang mit Verletzungen des Urheberrechts entwickelt (Tschohl 2016). Die Vertreter der Internetwirtschaft schlagen die Einrichtung einer unabhängigen Stelle mit richterlichen Kompetenzen vor, um nicht alleine die Verantwortung für Entscheidungen über Netzsperren tragen zu müssen. Der Vorschlag basiert auf der Idee einer Gewährleistung eines Interessensausgleichs zwischen Urheberrecht, Informationsfreiheit und Meinungsfreiheit. In institutioneller Hinsicht wird dafür ein Kollegialorgan mit richterlichem Einschlag, die Telekom-Control-Kommission (TKK), als „Clearingstelle“ sowie eine „sukzessive Kompetenz“ vorgeschlagen, bei der im Beschwerdefall die Entscheidung der Behörde außer Kraft tritt und die Zivilgerichte zuständig werden.

Bereits die wenigen Beispiele verdeutlichen die Bandbreite möglicher Institutionalisierungen von Gremien, die sich mit Nutzerbeschwerden und Berufungsverfahren auf nationaler Ebene auseinandersetzen: Die Bandbreite reicht von monothematischen single-issue Organisationen (z.B. zum Thema Urheberschutz) über Einrichtungen zum Umgang mit illegalen Inhalten im engen Sinn, bis hin zu Organisationen, die für sämtliche Formen von Inhalten und Problemstellungen im Bereich der Internetkommunikation zuständig sein könnten.

Die Form der Institutionalisierung und der Grad der staatlichen Involvierung reichen von einer Selbstregulierung im weiten Sinn mit breiter Stakeholdereinbindung über Ko-Regulierung unter staatlicher Aufsicht bis hin zu unabhängigen Regulierungsbehörden mit hoheitlichen Befugnissen.

Zentrale Fragen für die Einrichtung einer Ombudsstelle betreffen deren Aufgabenbereich, sowie die Berechtigung und den Modus Fälle vorzubringen. Denn der Aufgabenbereich und die Antragsmodalitäten wirken sich direkt auf die Anzahl der Fälle aus, die von einer Ombudsstelle bearbeitet werden müssen (*Mengengerüst*) und bestimmen somit auch über den Aufwand und den Ressourcenbedarf (z.B. Anzahl der Mitarbeiter, Verfahrensablauf, Frequenz der Entscheidungssitzungen, etc.). Eine Ombudsstelle muss vor Missbrauch und einer Überflutung mit Fällen geschützt sein. Wichtig ist deshalb, dass eine Ombudsstelle nicht als erste Anlaufstelle für die Meldung problematischer Inhalte fungieren soll, sondern für Berufungen gegen die Moderationsentscheidungen von Plattformen die als ungerechtfertigt empfunden werden. Ombudsstellen sollen die Arbeit von Meldestellen und Inhaltsmoderation durch Plattformen nicht ersetzen, sondern für umstrittene Fälle ergänzen.

Ein weiterer wichtiger Aspekt der Institutionalisierung von Berufungsgremien ist die Frage welchen Grad der *Verbindlichkeit* Ergebnisse aus nationalen Berufungsverfahren für die internationalen Plattformen hätten: Verbindliche Entscheidungen oder lediglich Empfehlungen. Denkbar wäre, dass sich analog zu den „vertrauenswürdigen Berichterstattern“ (Meldestellen) auch „vertrauenswürdige Berufungsgremien“ (Ombudsstellen) etablieren, deren Beratungsergebnisse von Plattformen aufgrund hoher Input-Legitimität durch breite Stakeholdereinbindung und umsichtiger Interessenabwägungen akzeptiert werden.

Von hoher Bedeutung für die Institutionalisierung ist auch das Verhältnis der Berufungsgremien zu den Internet-Plattformen. Grundsätzlich sollen die Berufungsgremien von den Plattformen unabhängig agieren. Sie sind jedoch für ihre Verfahren auf die Kooperation, den Informationsaustausch und die Umsetzung von Entscheidungen oder Empfehlungen durch die Plattformen angewiesen. Sofern Plattformen mit den Berufungsgremien *freiwillig* kooperieren, notwendige Informationen übermitteln und Ergebnisse (z.B. als Empfehlungen) beachten, könnten die Berufungsgremien im Sinne von „zivilgesellschaftlichen Ombuds-/Schlichtungsstellen" betrieben werden. Wenn die freiwillige Kooperation scheitert, müsste eine Stelle mit adäquaten rechtlichen Grundlagen und Befugnissen ausgestattet werden, um ihre Funktionen sinnvoll ausüben zu können. Mit steigender staatlicher/rechtlicher Involvierung wachsen jedoch auch Skepsis und Druck in Hinblick auf mögliche Beschränkungen der Meinungsfreiheit durch den Staat.

Die Einrichtung von Ombudsstellen bietet viele Potenziale für die Behandlung von Einzelfällen und einem Fokus auf Konsumentenschutz. Sie sorgt jedoch nicht für eine generelle Kontrolle der Normen und Verfahren der Inhaltsregulierung.

4.2.3 Institutionen für Aufsicht und Kontrolle

Normen, Prozesse/Verfahren und Ergebnisse der Inhaltsregulierung sollen Gegenstand von regelmäßigen Überprüfungen sein. Mit regelmäßigen Überprüfungen soll festgestellt werden, ob Normen adäquat sind, ob prozedurale Standards eingehalten werden und welche Ergebnisse mittels Inhaltsmoderation erzielt werden. Expertinnen und Experten sind sich darüber einig, dass solche Überprüfungen von unabhängigen externen Einrichtungen durchgeführt werden sollten, die sowohl von den Plattformen als auch von Politik und Regierung *unabhängig* sind. Aktuelle Vorschläge für die Institutionalisierung von Kontrolle und Aufsicht divergieren auf dem Kontinuum zwischen experten-basierten Auditing-Systemen bis hin zu einer formalen Aufsicht durch unabhängige Regulierungsbehörden, die dafür mit formalen Auskunftsrechten („data gathering power") ausgestattet werden müssten, um die Moderationspraxis kontrollieren zu können.

Externe Kontrolle durch unabhängige Prüfstellen („auditing")

Im Rahmen der Kontrolle geht es u.a. um die Frage wie die Inhaltsregulierung in der Praxis funktioniert, ob Zielsetzungen erreicht werden und ob Fehlfunktionen wie overblocking oder systematische Verzerrungen (Bias) vorliegen.

Für entsprechende Überprüfungen (Audits) und die Verbesserung der Transparenz besteht erheblicher Handlungsspielraum und Reformbedarf. Praktisch berichteten Plattformen über viele Jahre hinweg lediglich die Anzahl an takedowns in Folge rechtlicher Interventionen. Über Löschungen auf Grundlage der eignen Community Regeln gab es lange keine Transpa-

renz. 2018 haben YouTube und Facebook im Rahmen von Transparency Reports weiterreichende Informationen zur Moderationspraxis veröffentlicht. In Deutschland wurde mit dem NetzDG eine Berichtspflicht eingeführt. Gleichzeitig wurden Zugangsmöglichkeiten für wissenschaftliche Forschung als Konsequenz aus dem Cambridge Analytica Skandal beschränkt. Stattdessen wurden von Facebook gezielt Kooperationen mit ausgewählten wissenschaftlichen Einrichtungen im Rahmen des Social Science One Programms mit dem Social Science Research Council (SSRC) initiiert. Im Rahmen freiwilliger Kooperationen gibt Facebook seinen Partnern aus der Forschung beschränkten, stark kontrollierten Zugang zu Systemen und Daten und nimmt zudem an Überprüfungen zu ausgewählten Problemkomplexen teil, die jedoch v.a. auf die USA bezogen sind, wie z.B. zur Frage ob Facebook einen Bias gegen konservative US-Politiker aufweist (Ash et al. 2019, 19). Trotz punktueller Initiativen in Richtung höherer Transparenz und Kontrolle ist es für externe Einrichtungen bis dato jedoch weiterhin kaum möglich die Moderationspraxis der Plattformen und ihre potenziell problematischen Implikationen zu überprüfen.[114] Nicht zuletzt deshalb gibt es zu vielen Entwicklungen auf den Plattformen (Desinformation, Filterblasen, Manipulation) viele Spekulationen aber wenig empirische Evidenz, die sich als solide Grundlage für die Entwicklung von Reformen eignen würde.

„Derzeit ist es fast unmöglich, systematische Untersuchungen über die Verbreitungswege, die beteiligten Akteure und die Reichweite von Desinformationskampagnen innerhalb der Social-Media-Plattformen anzustellen. Wir können derzeit kaum systematisch untersuchen, wie viel Desinformation sich auf welchen Social-Media-Plattformen befinden, wie sehr das Design der Plattformen zur Verbreitung von Falschinformationen beiträgt, wer Desinformation systematisch verbreitet oder wie stark der Einfluss auf unsere politischen Debatten ist." (Sängerlaub 2019, 2)

Effektives Auditing setzt Zugang zu Systemen und Daten sowie Know-how für systematische Analysen voraus. Dafür bestehen in der Praxis vielfältige Hindernisse. Plattformen haben grundsätzlich kein Eigeninteresse an Kontrolle von außen und es ist wenig wahrscheinlich, dass sie sich freiwillig umfassenden Audits durch eine Vielzahl unabhängiger Dritte unterwerfen. Eine weitere Hürde für Auditing und Kontrolle dürfte ein Mangel an Expertise und Know-How sein, um gute Audits durchzuführen. Speziell für staatliche Behörden ist Plattformkommunikation ein neues Terrain, das für Kontrolle und Auditing eine Kooperation mit Expertinnen und Experten aus Wissenschaft und Wirtschaft erfordert. Zudem bestehen bei direkten staatlichen Kontrollen erhebliche Bedenken hinsichtlich interessengeleiteter Eingriffe in die Meinungsfreiheit.

Eine der Optionen zur Kontrolle sind deshalb *unabhängige Prüfstellen für Auditing* durch unabhängige Expertinnen und Experten. Diese sind vor allem geeignet, wenn es um die Überprüfung von technischen Anwendungen, ihrer Funktionsweise und ihren Auswirkungen geht. Gillespie (2018a, 214f.) und auch Ash et al. (2019, 19) schlagen zudem Beratungspanels (Expert

[114] Für einen Überblick zu vielfältigen Beschränkungen für externe Überprüfungen siehe Sängerlaub (2019, 28).

Advisory Panel; Content Policy Advisory Group) für Plattformen vor, denen neben Expertinnen und Experten auch Regulatoren, Wissenschaft sowie Vertreterinnen und Vertreter zivilgesellschaftlicher Gruppen angehören. Zusätzlich und noch weiter greifend könnten technische Schnittstellen geschaffen werden, die einen Zugriff von außen erlauben und dezentrale Audits ermöglichen, zum Beispiel für wissenschaftliche Expertinnen und Journalistinnen (Gillespie 2018, xv; Sängerlaub 2019, 27f.).

Die unterschiedlichen Formen von Auditing-Einrichtungen haben jeweils spezifische Vor- und Nachteile. Eine der zentralen Fragen dabei ist, welche Institutionen und Akteure im Rahmen von Audits Zugang zu Plattformen und ihren Moderationsverfahren und -daten erhalten sollen. Vertrauen ist ein wichtiger Faktor, denn Plattformen müssen Zugang zu Moderationsprozessen und -daten gewähren. Daten können für die breite Öffentlichkeit oder nur für zuständige Regulatoren und Prüfstellen oder für ausgewählte zivilgesellschaftliche Expertinnen und Experten wie wissenschaftlichen Einrichtungen zur Verfügung gestellt werden. Zu breiter Zugang eröffnet Risiken hinsichtlich Manipulation und Imitation. Daher wird mitunter empfohlen, einen beschränkten Zugang für Experten-Panels mit formalen Kontrollfunktionen zu gewähren (Ash et al. 2019, 19). Diese können die Inhaltsregulierung überprüfen, ohne die Codes und die genaue Funktionsweise der Kuratierung und Moderation von Inhalten der breiten Öffentlichkeit offenzulegen (Gillespie 2018, xv).

Externe Kontrolle durch unabhängige Aufsichtsinstanzen („oversight“)

Für die Institutionalisierung der Kontrolle werden auch formale Aufsichtsinstanzen vorgeschlagen. Solche Vorschläge reichen von einer „light touch“ Aufsicht in Form von internationaler Selbstregulierung bis zu nationalen unabhängigen Behörden mit mehr oder weniger weitreichenden Aufsichtskompetenzen. Konkrete Vorschläge wurden zuletzt vor allem im Vereinigten Königreich entwickelt und unterbreitet.

Global Partners Digital (vgl. Bradley & Wingfield 2018) schlagen beispielsweise ein Aufsichtsmodell für die internationale Ebene vor. Sie plädieren für die Einsetzung einer internationalen Aufsichtseinrichtung (OPSO, Independent Online Platform Standards Oversight Body), die als *Multistakeholdergremium* institutionalisiert ist und durch die Plattformen finanziert wird. Plattformen sollen unter Einbindung von Expertinnen und Experten und relevanten Stakeholdern *Richtlinien* (Online Platform Standards) entwickeln. Das OPSO soll die Einhaltung der Richtlinien durch die Plattformen überwachen und periodische Berichte mit Empfehlungen veröffentlichen. Bradley & Wingfield (2018, 27) favorisieren eine internationale Aufsicht, weil nationale Regelungen für die globalen Plattformen unangemessen sind und die Plattformen zwingen eine Reihe unterschiedlicher regionaler Anforderungen zu befolgen, obwohl die Herausforderungen globaler Natur sind. Darüber hinaus würden nationale Regulierungen aufgrund schwacher Grundrechtsstandards in vielen Ländern zu Zensur führen. Nationale Regulierungseinrichtungen haben in diesem Modell keine dezidierte Funktion.

Die Grundrechtsorganisation Artikel 19 schlägt vor, eine Aufsichtsinstanz als „Social Media Council“ im Rahmen einer *Selbstregulierung* auf internationaler oder nationaler Ebene einzurichten (vgl. Article 19 2018, 21f.). Um die Unabhängigkeit zu sichern, sollten dem Council neben den Plattformen selbst auch Repräsentanten aller relevanten Stakeholder angehören, darunter Medienverbände, Regulierungsbehörden, Interessengruppen im Bereich Grundrechte und Meinungsfreiheit, Wissenschaft und Vertreter der Zivilgesellschaft. Das Council soll generelle Aufsichtspflichten gegenüber Plattformbetreibern wahrnehmen, und z.B. *Verhaltens-/Ethikkodizes* entwickeln und deren Einhaltung überprüfen. Als Vorbild fungieren die unabhängigen Presseräte, die traditionell als Selbstregulierung etabliert sind und über berufsethische Regeln im Journalismus entscheiden. Das Council soll entsprechend den Vorschlägen von Article 19 zudem operativ als Beschwerde- bzw. Berufungsstelle gegenüber Entscheidungen der Plattformbetreiber im Rahmen der Inhaltsmoderation fungieren.

Bunting (2018, 25f.) schlägt einen „Oversight Body“ vor, der auf nationaler Ebene als unabhängige Aufsichtsinstanz eingerichtet werden soll. Die Aufgaben des Aufsichtsorgans umfassen u.a. die Entwicklung, Promotion und Durchsetzung eines *Verhaltenskodex* für Inhaltsmoderation durch Plattformbetreiber. Die Grundlagen für das Aufsichtsorgan sollen durch einen rechtlichen Rahmen festgelegt werden, der den Wirkungsbereich des Aufsichtsorgans durch die Definition von „Online Content Intermediären“ absteckt, und mit dem die wesentlichen Zuständigkeitsbereiche definiert sind, v.a. Zuständigkeiten zur Überprüfung von Risiken, zur Entwicklung und den Zielen eines Verhaltenskodex, sowie zur Setzung von Anreizen und Sanktionen um Vorgaben des Verhaltenskodex durchzusetzen. Die Institutionalisierung könnte laut Bunting (2018, 28) als *industrie-geleitete Ko-Regulierung* nach Vorbild der Advertising Standards Authority (ASA) im Vereinigten Königreich erfolgen. Dabei wird die praktische Kontrolle – innerhalb eines rechtlichen Rahmens – durch die Industrie und unter Einbindung wesentlicher Stakeholder durchgeführt. Staatliche Einrichtungen setzen den rechtlichen Rahmen, führen Evaluierungen durch und greifen ein, wenn die Regulierung durch die Industrie nicht zu Stande kommt oder versagt. Im Fall eines Versagens soll eine *behördliche Aufsicht* installiert werden, entweder im Rahmen der bestehenden Aufsichtsorgane wie Ofcom oder in Form einer neuen Aufsichtsbehörde (Bunting 2018, 33).

Die LSE Commission on Truth, Trust and Technology (2018, 3 und 36f.)[115] empfiehlt die Einrichtung einer *Independent Platform Agency (IPA)*, die aus einer Steuer auf die Umsätze (digital platform levy) der Plattformbetreiber finanziert werden soll. Die IPA soll *per Gesetz eingerichtet* werden, aber von der Regierung unabhängig agieren und nur dem Parlament gegenüber verantwortlich sein. Die Agency soll sich zunächst vor allem Problemstellungen widmen, die mit Desinformation in Verbindung stehen. Sie soll jedoch selbst keine direkten regulatorischen Entscheidungen über Inhalte treffen, sondern eine *Beobachtungs- und Beratungsfunktion* ausüben. Dazu zählen kurzfristig die Analyse und Berichterstattung über Trends im Bereich der

[115] Für den Vorschlag einer Regulierungsbehörde siehe auch Tambini (2018).

Nachrichten- und Informationsverbreitung auf Plattformen, die Beobachtung und Evaluierung der Selbstregulierungsinitiativen der Plattformen hinsichtlich Qualitätskontrolle und Signalisierung von Vertrauenswürdigkeit sowie Inhaltsfilterung; die Koordination von Initiativen und Programmen zur Verbesserung der Medienkompetenz; und die Beteiligung an der Kontrolle der Internet-Plattformen gemeinsam mit anderen kompetenten und zuständigen Regulierungsbehörden, z.B. hinsichtlich der Überprüfung der Marktdominanz der Plattformen. Zur Ausübung ihrer Beobachtungsfunktion soll der Platform Agency vor allem das Recht eingeräumt werden, alle relevanten Informationen und Daten von den Plattformbetreibern einzuholen, sowie Sanktionsmöglichkeiten (Geldbußen), wenn die Herausgabe der Informationen verweigert wird. Mittel- und langfristig soll die Agency in Zusammenarbeit mit dem Parlament und der Zivilgesellschaft auch Standards für Plattformen vorgeben können, wenn die Selbstregulierung der Plattformen versagt und bei Bedarf selbst Regulierungsfunktionen ausüben.

Auch zur Frage der Aufsicht und Kontrolle verdeutlichen die wenigen Beispiele die Bandbreite möglicher Institutionalisierungen. Für die konkrete institutionelle Ausgestaltung solcher Einrichtungen stellen sich etliche Herausforderungen, z.B. hinsichtlich der Ansiedlung auf nationaler oder internationaler Ebene, dem Grad einer staatlichen Involvierung, den gegebenenfalls notwendigen rechtlichen Grundlagen, ihrer Zusammensetzung, ihres Auftrags und den Kompetenzen zur Durchsetzung von Standards.

Zur Aufsicht vorgeschlagen werden sowohl nationale als auch internationale Instanzen. Internationale Einrichtungen stoßen vermutlich auf größere Akzeptanz seitens der Plattformbetreiber, erreichen aber oftmals nur sehr generelle Mindeststandards und sind zumeist durchsetzungsschwach. Nationale Einrichtungen sind leichter zu initiieren und können mit stärkeren Befugnissen ausgestattet werden. Sie bergen allerdings das Risiko, dass zu viele, zu unterschiedliche nationale Vorgaben Plattformen vor kaum bewältigbare Herausforderungen stellen und Innovationen bremsen.

Einigkeit besteht darüber, dass als normative Bezugsdokumente für die Inhaltsregulierung nicht vorrangig rechtlichen Regeln,[116] sondern vor allem Richtlinien in Form von Verhaltenskodizes oder Plattformstandards gelten sollen, die von der Industrie in Zusammenarbeit mit Stakeholdern entwickelt werden. Analog zur Frage nach der Ansiedlung einer Aufsichtsinstanz stellt sich dabei die Frage, ob solche Verhaltenskodizes auf nationaler oder internationaler Ebene entwickelt und implementiert werden sollen, bzw. ob verschiedene nationale Verhaltenskodizes oder ein internationaler Verhaltenskodex angemessen ist.

Einigkeit besteht darüber, dass die Instanzen „staatsfern“ als unabhängige Aufsichtsorgane eingerichtet werden sollen. Für die Form der Institutionalisierung und den Grad der staatlichen Involvierung in der Aufsicht werden jedoch verschiedene Vorschläge eingebracht. Diese

[116] Eine Ausnahme bildet hier Deutschland, wo mit dem NetzDG u.a. verbindliche Fristen für die Entfernung rechtswidriger Inhalte vorgegeben wurden.

reichen von einer Selbstregulierung mit breiter Stakeholdereinbindung über Ko-Regulierung innerhalb eines rechtlichen Rahmens bis hin zu unabhängigen Regulierungsbehörden mit hoheitlichen Befugnissen. Manche Vorschläge (Bunting 2018), sehen ein dynamisches Modell vor, indem eine staatliche Aufsicht dann eingeführt wird, wenn Selbstregulierung versagt und ein Kodex von der Behörde erstellt wird, wenn die Industrie keinen eignen Kodex vorlegt.[117]

Hinsichtlich der Aufträge für die Aufsichtsinstanzen besteht Einigkeit darüber, dass diese die Einhaltung der Vorgaben der Verhaltenskodizes kontrollieren, regelmäßiges Monitoring betreiben und die Öffentlichkeit über die Umsetzung der Vorgaben informieren sollen. Manche Vorschläge gehen über solche Basisfunktionen hinaus und umfassen zusätzliche Funktionen wie die Koordination von Medienkompetenzprogrammen (z.B. LSE 2018). Wichtig erscheint jedenfalls die Funktion die Inhaltsregulierung durch Plattformen regelmäßig zu evaluieren. Bislang fehlt es an länderspezifischen Informationen und Daten zur Anzahl der Meldungen nach Meldequellen, zu den Reaktionen der Plattformen im Rahmen der Moderation sowie an Angaben über den Umfang, Art und Reaktionszeit der gesetzten Maßnahmen. Aufsicht sollte Transparenz über die Inhaltsmoderation herstellen und über Instrumente verfügen, um die relevanten Informationen von Plattformen einzuholen, z.B. mit Hilfe verpflichtende Transparenzberichte auf Basis standardisierter Kriterien über die Inhaltsregulierung.

Weniger klar ist über welche Durchsetzungsinstrumente die Aufsichtsinstanzen verfügen sollen. Etliche „light touch" Vorschläge sehen keine Durchsetzungsinstrumente vor, oder bleiben hinsichtlich der Durchsetzung vage. Zum Teil sehen die Vorschläge Möglichkeiten vor, Anreize und Sanktionen zu setzen, um Vorgaben aus einem Verhaltenskodex durchzusetzen. Wenn freiwillige Zusammenarbeit nicht zustande kommt erscheinen formale Auskunftsrechte für die Aufsichtsinstanzen und Auskunftsverpflichtungen für die Plattformen zentral. Sanktionsmöglichkeiten (Geldbußen) werden beispielsweise vorgeschlagen, wenn die Herausgabe von Informationen durch die Plattformen verweigert wird (LSE 2018).[118]

Die Etablierung von Aufsichtsinstanzen und die Umsetzung von Aufsichtsfunktionen erfordert personelle und finanzielle Ressourcen. Hohe Kompetenzen und einen hohen Ressourceneinsatz sind insbesondere erforderlich, wenn die Instanzen nicht nur eine formale Rechtsaufsichtsfunktionen wahrnehmen, sondern selbst aktiv Trends in der Plattformkommunikation beobachten hinsichtlich riskanter Entwicklungen prüfen sollen (Monitoringfunktion). Damit stellt sich die Frage welche Einrichtungen über die dafür notwendigen Ressourcen (Personal, Know-how) verfügen bzw. wie diese Aufgaben finanziert werden sollen. Die Vorschläge zur Finanzierung umfassen in der Regel Finanzierungsbeiträge der Plattformunternehmen. Die

[117] Vgl. Bunting (2018, 22): „An alternative approach would be to empower a regulator to provide for a more comprehensive, but higher-level Code of Practice, defining both broad content standards and procedural principles, but with fewer specific requirements".

[118] Im Themenkomplex Information und Auskunft sind darüber hinaus mögliche Beschränkungen durch datenschutzrechtliche Vorgaben beim grenzüberschreitenden Datentransfer zu berücksichtigen.

institutionellen Voraussetzungen für die Einrichtung von Aufsichtsinstanzen hinsichtlich Personal, Kompetenz und Know-How sind von Land zu Land unterschiedlich. Während beispielsweise im Vereinigten Königreich aktuell sowohl die Einrichtung einer neuen Aufsichtsbehörde als auch einer Erweiterung der Kompetenzen für die bestehende Kommunikationsbehörde Ofcom diskutiert werden, wurden für Österreich noch keine konkreten Vorschläge zur Diskussion gestellt.

4.3 Fazit und Schlussfolgerungen zu Grundsätzen und Institutionalisierungsoptionen

Ein verantwortungsorientiertes Governance-Arrangement für Inhaltsregulierung auf Internet-Plattformen erfordert Grundsätze und Kernelemente für die institutionelle Ausgestaltung. Mit Hilfe von Literaturanalysen wurden verschiedene Vorschläge in der Literatur identifiziert.

Als Rahmenkonzept und *Leitprinzip bietet sich Verantwortlichkeit (Accountability)* an, verstanden als Bereitschaft und Verpflichtung, Verantwortung zu übernehmen, verantwortungsvoll zu handeln, Rede und Antwort zu stehen, Rechenschaft abzulegen und Konsequenzen zu tragen. So beschreibt Verantwortlichkeit einen Mechanismus, der über die organisatorische Festlegung von Zuständigkeits- und Verantwortungsbereichen und rechtliche Verursachungs- und Haftungsfragen hinausgeht.

Für Grundsätze eines Governance-Arrangements werden in der Literatur vielfältige Vorschläge eingebracht. Als zentrale Herausforderung gilt dabei, Verantwortlichkeit für Inhaltsregulierung zu fördern, ohne dadurch die Meinungsfreiheit zu gefährden.

Neben Verantwortlichkeit und Meinungsfreiheit zählen zu wichtigsten Grundsätzen:

- Das Primat der Plattformen bei der operativen Tätigkeit der Inhaltsregulierung
- Legitimität der Eingriffe und Verhältnismäßigkeit der eingesetzten Instrumente
- Transparenz hinsichtlich Normen und Verfahren der Inhaltsregulierung
- Transparenz und Information gegenüber Nutzern bei Einzelfallentscheidungen
- Faire Verfahren und Beeinspruchungsmöglichkeiten
- Transparenz und Überprüfbarkeit durch Offenlegung von Moderationsdaten
- Unabhängige Kontrolle, Überprüfung und Evaluierung der Inhaltsregulierung
- Stakeholder-Einbindung und Interessenabwägung bei Entscheidungen
- Demokratische Ansätze und Nutzereinbindung

Hinzu kommen Überlegungen und Vorschläge für Grundsätze, die weniger die Inhaltsregulierung im engen Sinn betreffen, sondern den Governance-Ansatz im Allgemeinen. Zu diesen Vorschlägen zählen:

- „Adaptive Governance Ansätze“, die laufende Überprüfungen und Reformprozesse ermöglichen, weil sich Nutzungsverhalten, Erwartungen und Dienstangebote wandeln.
- „Differenzierte Governance Ansätze“ die Unterschiede in den Reichweiten, Formaten, Geschäftsmodellen und den daraus resultierenden Risikopotenzialen berücksichtigen, wenn regulatorische Anforderungen formuliert und durchgesetzt werden.
- „Prozedurale Governance Ansätze“, die grundsätzliche Ziele und Standards für den Regulierungsprozess vorgeben, aber keine detaillierten Vorschriften hinsichtlich Verhalten und Instrumenten.

Als weitere zentrale Grundsätze für Governance-Ansätze gelten *Subsidiarität und Kontextsensibilität*. Internet-Plattformen agieren global, aber ihre Entscheidungen und Programme werden regional wirksam. Vor allem weil die Inhaltsmoderation der Plattformen Auswirkungen auf nationale Öffentlichkeiten hat, müssen die Plattformen sich auch diesen nationalen Gesellschaften und Öffentlichkeiten gegenüber „verantworten“, also Rede- und Antwort stehen und Rechenschaft ablegen. Daraus lässt sich die Forderung nach einer nationalen Rückkoppelung des Governance-Arrangements ableiten. Es ist wichtig eine stärkere Kontextsensibilität der Inhaltsregulierung zu erreichen, und Arrangements zu fördern, die auf nationale Besonderheiten Rücksicht nehmen, ohne dabei die Funktionsfähigkeit der Plattformen zu gefährden.

Deshalb wird in einem weiteren Analyseschritt untersucht, welche institutionellen Kernkomponenten so eine nationale Rückkoppelung der Inhaltsregulierung ermöglichen würden (Optionen).

Zu den identifizierten Möglichkeiten zählen folgende Institutionen für verantwortungsorientierte Kommunikation, für Melde- und Beschwerdemanagement sowie für Aufsicht und Kontrolle:

- Kommunikative Verantwortlichkeit durch nationale Plattformsprecher
- Knotenpunkt zu nationalen Öffentlichkeiten durch nationale Melde- und Beratungsstellen
- Kontextsensibilität durch Moderationsteams mit regionalem/nationalem Kontextwissen
- Geteilte Verantwortlichkeit durch unabhängige Berufungsverfahren
- Rechtstaatlichkeit und Machtverteilung durch Berufungsgremien auf nationaler Ebene
- Externe Kontrolle durch unabhängige Prüfstellen
- Externe Kontrolle durch unabhängige Aufsichtsinstanzen

Die Analyse verdeutlicht, dass institutionelle Optionen für eine nationale Rückkoppelung der Inhaltsregulierung auf Internet-Plattformen verfügbar sind. Die Bandbreite möglicher Institutionalisierungen reicht dabei von einer Selbstregulierung im weiten Sinn mit breiter Stakehol-

dereinbindung über Ko-Regulierung unter staatlicher Aufsicht bis hin zu unabhängigen Regulierungsbehörden. Für die konkrete institutionelle Ausgestaltung solcher Einrichtungen stellen sich etliche Herausforderungen, z.B. hinsichtlich der Ansiedlung auf nationaler oder internationaler Ebene, dem Grad einer staatlichen Involvierung, den gegebenenfalls notwendigen rechtlichen Grundlagen, ihrer Zusammensetzung, ihres Auftrags und den Kompetenzen zur Durchsetzung von Standards. Einigkeit besteht darüber, dass Aufsicht und Kontrolle sowohl von den Plattformen als auch von Politik und Regierung unabhängig organisiert sein müssen. Aktuelle Vorschläge für Kontrolle und Aufsicht reichen von expertenbasierten Auditingsystemen bis hin zu einer formalen Aufsicht durch unabhängige Regulierungsbehörden.

5. Reformoptionen für Österreich

Im fünften und letzten Abschnitt wird untersucht, welche Reformschritte notwendig wären, um Komponenten eines verantwortungsorientierten Governance-Arrangements in Österreich zu implementieren. Dabei geht es nicht um Details einer rechtlichen Ausgestaltung von Verantwortung, Haftung, Rechtsdurchsetzung und Rechtsschutz im komplexen Themenbereich der Regulierung von Internet-Inhalten. Die Analyse fokussiert auf die Identifikation von geeigneten institutionellen Komponenten eines verantwortungsorientierten Governance-Arrangements auf nationaler Ebene in Österreich und umfasst die Identifikation des Handlungsbedarfs für die Plattformbetreiber (z.B. Etablierung von Auskunftspersonen), Vorschläge für die institutionelle Weiterentwicklung des Melde- und Beschwerdemanagements (z.B. Ombudsstellen) und Überlegungen zur Etablierung von Aufsichtsfunktionen im nationalen Rahmen. Bei den im Folgenden vorgestellten Reformoptionen handelt sich um eine Zusammenstellung auf Basis von Literaturanalysen und theoretischen Überlegungen, die zur Validierung im Rahmen von Interviews mit österreichischen Expertinnen und Experten[119] hinsichtlich ihrer praktischen Relevanz, ihrer Umsetzbarkeit und möglicher Umsetzungshürden diskutiert wurden.

5.1 Rahmenbedingungen & -strategien

Wie für viele andere Regulierungsfragen im Internet gibt es auch zur Frage der Inhaltsregulierung auf Internet-Plattformen *keine "one-size-fits-all" Lösungen*. Eine adäquate Inhaltsregulierung erfolgt nicht alleine durch den Markt, die Plattformbetreiber, eine unabhängige Regulierungsbehörde oder den Staat. Es besteht der Bedarf für ein Zusammenspiel zwischen den verschiedenen Governance-Akteuren und eine problemadäquate Kombination von Maßnahmen auf verschiedenen Governance-Ebenen, die sich gegenseitig ergänzen und ineinander

[119] Die Gespräche mit Expertinnen und Experten fanden im Zeitraum Februar bis Mai 2019 statt. Interviewpartnerinnen und -partner waren Maximilian Schubert (ISPA), Bernhard Jungwirth und Karl Gladt (ÖIAT, Internet Ombudsmann), Caroline Kerschbaumer und Hans Dieter Schindlauer (ZARA), Ingrid Brodnig, Klaudia Zotzmann-Koch, Maria Windhager und Ben Wagner. Die Vorschläge in Abschnitt 5 müssen nicht den Meinungen der Interviewpartnerinnen und -partner entsprechen.

greifen (Komplementarität). Idealtypisch können Maßnahmen auf einer der Governance-Ebenen eine Hebelwirkung entfalten und Maßnahmen auf anderen Ebenen bewirken.

Im Folgenden werden vor allem Optionen identifiziert, um ein verantwortungsorientiertes Governance-Arrangement in Österreich zu implementieren. Eine etwaige Umsetzung muss beim Thema Plattformregulierung internationale Kontextbedingungen mitberücksichtigen. Insgesamt muss eine *kohärente Governance-Architektur* entwickelt werden, die nationale und internationale Instanzen und Normen widerspruchsfrei verbindet. Nationale Initiativen können dazu dienen, Strategien auszuprobieren und europäische Lösungen anzustoßen. Für Governance-Arrangements mit staatlichen und privaten Beiträgen ist zudem darauf zu achten, dass Kompetenzen und Verantwortlichkeiten klar zugeordnet sind (Saurwein 2011, 352).

Die Analyse von Kontextfaktoren für Inhaltsregulierung legt ein Regulierungsarrangement nahe, das in der operativen Praxis von den Plattformen selbst betrieben wird und gleichzeitig unter öffentlicher Kontrolle steht (Abschnitt 2). Bei einem Mix aus privater Umsetzung und öffentlicher Kontrolle bestehen Möglichkeiten, um Kontrollsysteme zu entwickeln, die auf Kooperation zwischen Plattformen und öffentlichen Aufsichtsinstanzen basieren. Wenn kooperative Ansätze scheitern, muss Kontrolle mit Hilfe rechtlicher Maßnahmen durchgesetzt werden (vgl. Ayres & Braithwaite 1992, „Enforcement Pyramide“).

Die Reform des institutionellen Arrangements der Inhaltsregulierung im Internet und auf Internet-Plattformen wird ein iterativer Prozess mit laufenden Anpassungen im Sinn einer *adaptiven Governance* sein. Auch im Rahmen dieser Studie soll kein neues, finales Arrangement entwickelt werden, sondern es sollen einige Reformoptionen skizziert werden, die sich dazu eignen die Inhaltsregulierung auf Internet-Plattformen national zu koppeln, weil seitens der Plattformen eine Verantwortlichkeit gegenüber nationalen Öffentlichkeiten besteht. Solche Ansatzpunkte können geprüft, diskutiert, ggf. implementiert und im Sinne einer adaptiven Governance weiterentwickelt werden.

5.2 Institutionelle Reformoptionen

Kommunikative Verantwortlichkeit: Bevollmächtigte nationale Auskunftsverantwortliche

Accountability kann als Mechanismus verstanden werden, der auf Rede-und-Antwort-Stehen abzielt und damit eine starke kommunikative Komponente beinhaltet. Auch Internet-Plattformen sollten ihr Handeln stärker erklären und rechtfertigen, und sich einer kritischen nationalen Öffentlichkeit und Diskussion stellen. Weil die Inhaltsmoderation der Plattformen Auswirkungen auf nationale Öffentlichkeiten hat, müssen die Plattformen sich auch vor diesen nationalen Gesellschaften und Öffentlichkeiten gegenüber „verantworten“ indem sie ihr Handeln erklären und rechtfertigen.

Die Einrichtung von Instanzen für verantwortungsorientierte Kommunikation bei den Internet-Plattformen ist aufgrund des Einflusses der Plattformen, den vielfältigen Skandalen und

offenen Fragen überfällig. Vorgeschlagen wird deshalb die Installierung einer auskunftsberechtigten und -bevollmächtigten Person (eines Sprechers bzw. einer Sprecherin) auf nationaler Ebene in Österreich die auf Kritik reagieren, Rede- und Antwort stehen, und sich in nationale Debatten einbringen kann. Nationale Sprecher sollen kommunikative Aufgaben erfüllen die über die haftungstechnisch relevante Funktion (z.B. eines Zustellungsbevollmächtigten) hinausgehen. So kann kommunikative Verantwortung gegenüber nationalen Öffentlichkeiten besser wahrgenommen werden. Dadurch entsteht ein Beitrag zu einer „Verantwortungskultur", die sich mit Jarren (vgl. 2018) nicht allein durch hoheitliche Maßnahmen institutionalisieren lässt, sondern sich nur durch Dialoge entwickeln kann. An diesen Dialogen sollen in Zukunft auch die Plattformen selbst aktiver teilnehmen.

Nationale Rückkoppelung im Beschwerdemanagement

Mit der steigenden Verbreitung und Nutzung des Internet wird die Anzahl illegaler und problematischer Inhalte zunehmen und es werden sich weiterhin viele Fälle in verschiedensten Arten von Grauzonen finden. Wenn die Anzahl an problematischen Inhalten und die Anzahl an Beanstandungen steigen, steigt auch der Bedarf an Institutionen für das Beschwerdemanagement. Dabei ist es wichtig, nicht alle diesbezüglichen Funktionen auf eine einzelne Institution zu konzentrieren, sondern in Formen geteilter Verantwortung in „kaskadischen Systemen" (Spielkamp 2018) und in unterschiedlichen Prozessphasen zu denken. Endgültige Moderationsentscheidungen sollen weiterhin von den Plattformen selbst getroffen werden. Die Verantwortung für vorgelagerte Melde- und nachgelagerte Berufungs- und Schlichtungsverfahren könnte jedoch von Plattformen stärker delegiert und ausgelagert werden. Dabei bieten sich Formen der Aufgaben- und Verantwortungsverteilung im Beschwerdemanagement an, in dem auch geeignete Einrichtungen auf nationaler Ebene eine Rolle spielen können.

Meldewesen: Förderung und Ausbau nationaler Meldestellen

Im Meldewesen haben sich in Österreich mittlerweile mehrere Anlaufstellen etabliert (Stopline, Beratungsstelle #GegenHassimNetz, BMI Meldestellen, Internetombudsmann) denen problematische Inhalte gemeldet werden können. Dem teils öffentlichen, teils zivilgesellschaftlich organisierten Meldewesen in Österreich wird im Allgemeinen eine effiziente Funktionsweise zu attestiert (Berka & Trappel 2019, 92), trotzdem bestehen spezifische Verbesserungspotenziale hinsichtlich Ressourcenausstattung, Professionalität und Kooperation.

Als Hindernis in Bezug auf die Zusammenarbeit werden die Kommunikationsstrukturen zwischen beteiligten Akteuren wie Meldestellen, Behörden und Plattformen genannt, beispielsweise Unklarheiten bezüglich der Kontaktaufnahme und der Form von Meldungen gegenüber Plattformen und Behörden (vgl. Berka & Trappel 2019, 55). Mittlerweile kooperieren die Plattformen verstärkt mit zivilgesellschaftlichen Meldestellen im Rahmen von Trusted Flagger-Programmen. Die verstärkte Zusammenarbeit mit Organisationen der Zivilgesellschaft hat zu stichhaltigeren Meldungen, einer wirksameren Behandlung und besseren Ergebnissen bei der Reaktion auf Meldungen durch die Plattformen geführt (Europäische Kommission 2017).

Wenn es um illegale und gefährliche Inhalte geht, so bilden nationale Meldestellen die wichtigen Seismographen und Knotenpunkte zwischen der Kommunikation in nationalen Öffentlichkeiten, und den Plattformen, über die diese Kommunikation verläuft. Speziell die Trusted Flagger-Programme – dem in Österreich z.B. ZARA angehört – schaffen einen direkten moderierten Beschwerdekanal von nationalen Öffentlichkeiten zu den Plattformen. Gerade bei problematischen Inhalten im Internet, die potenziell extrem schnell und weit verbreitet werden können, und zeitnahe Reaktionen erfordern, sind prompte und vertrauenswürdige Überprüfungen sowie direkte Kanäle im Beschwerdemanagement besonders wichtig. Sie können dazu beitragen rasch und professionell auf problematische Inhalte zu reagieren und Opfer zu schützen. In einem verantwortungsorientierten Governance-Arrangement spielen die bestehenden Meldestellen auf nationaler Ebene, die Inhalte schnell und professionell prüfen und vertrauenswürdige Meldungen direkt an die Plattformen richten, eine wichtige Rolle. Die strukturierte Zusammenarbeit der Meldestellen mit den Plattformen im Rahmen von Trusted Flagger-Programmen sollte weitergeführt und ausgebaut werden.

Wesentlich für das Funktionieren des Meldewesens ist eine *adäquate Mittelausstattung*. Nachdem die Meldestellen wichtige Aufgaben für die Öffentlichkeit erfüllen, sollte die Ressourcenausstattung in Österreich weiterhin mit öffentlichen Mitteln sichergestellt und ausgebaut werden.[120] Nachdem die Meldestellen durch ihre Vorprüfung auch Moderationsaufgaben der Plattformen unterstützen wäre auch eine finanzielle Beteiligung der Plattformen am Betrieb der Meldestellen gerechtfertigt. Ein hoher Bekanntheitsgrad ist notwendig damit Meldestellen ihre Aufgaben wahrnehmen können. Deshalb sollen Meldestellen über ausreichend Budget für Öffentlichkeitsarbeit verfügen, um ihre Funktionen zu bewerben und Bekanntheit zu steigern. Generell ist es wichtig, dass Meldefunktionen mit einem benutzerfreundlichen Design gestaltet sind, das an den Bedürfnissen der Menschen orientiert ist und auch von Laien leicht bedient werden kann.

Als Anlaufpunkte für Personen mit Beschwerden ist eine hohe Problemnähe gegeben, die auch (weiterhin) für *Beratungsfunktionen* genutzt und erweitert werden soll. Auch ein Ausbau von Meldestellen und ihrer Zuständigkeit entlang weiterer Problemstellungen und neuer Herausforderungen (z.B. Desinformation) könnte nützlich sein und geprüft werden.

Inhaltsmoderation: Moderatoren mit regionalem/nationalem Kontextwissen

Vielfach wird darauf verwiesen, dass Inhaltsmoderation und Löschentscheidungen zentral und automatisiert mit zu wenig Sensibilität für die jeweiligen nationalen Kontexte getroffen werden. Plattformen sollen deshalb Moderatorinnen und Moderatoren mit nationalem/regionalem Kontextwissen beschäftigen. Facebook ist dabei seine Moderationsteams stärker zu re-

[120] So erfolgte die Finanzierung der Einrichtung der Beratungsstelle von ZARA über das Digitalisierungsbudget des Staatssekretariats für Medien, Kunst und Kultur.

gionalisieren. Trotzdem erscheinen eigene Moderationsteams für kleine Länder wie Österreich kurzfristig praktisch wenig realistisch, doch zumindest sprach-/kulturregionale und gut betreute Moderationsteams wären sinnvoll.

In Deutschland wurden, nicht zuletzt unter dem Druck des NetzDG, eigene Moderationsteams aufgebaut. Möglicherweise könnten sich für Inhaltsmoderation auch länderübergreifende Synergien anbieten, indem zum Beispiel in deutschen Moderationsteams auch Mitarbeiter mit österreichischem Kontextwissen beschäftigt werden. Entsprechende Vorkehrungen im Rahmen der Rekrutierung sind durch Plattformbetreiber sicherzustellen.

Berufung & Streitbeilegung: Einrichtung einer nationalen Ombudsstelle

Plattformen treffen milliardenfach automatisierte oder manuelle Moderationsentscheidungen nach Unternehmensstandards. Gehen Plattformbetreiber außerhalb von rechtstaatlichen Verfahren gegen problematische Inhalte vor, so kann dies zu overblocking führen, die Meinungsfreiheit einschränken und gegen Grundsätze von Transparenz, Legitimität, Verhältnismäßigkeit und Verantwortlichkeit verstoßen. Umgekehrt wurden Internet-Plattformen auch lange dafür kritisiert, dass sie gegen eine Vielzahl an problematischen Inhalten keinerlei Maßnahmen zu setzen bereit waren. Egal ob Inhalte gelöscht werden oder sichtbar bleiben, für Beschwerdeführer, deren Reklamationen zurückgewiesen werden, ist der Verweis der Plattformbetreiber auf „Community-Standards" meistens wenig befriedigend. Und der Rechtsweg als Alternative nicht geeignet, zu aussichtlos, zu langwierig, zu kostspielig, und Staatsanwaltschaften sowie ordentliche Gerichte wären mit einer Vielzahl an Beschwerden ohnehin überlastetet. Bei etlichen Problemlagen stellt sich die Frage wie Moderationsentscheidungen der Plattformbetreiber beeinsprucht werden können, und welche Berufungsverfahren sich anbieten. Auch dabei könnten Instanzen auf nationaler Ebene eine Rolle spielen.

Dafür bietet sich die Einrichtung einer staatsfernen nationalen österreichischen *Ombuds- und Schlichtungsstelle für Inhaltsmoderation im Internet* an. Die Stelle selbst soll keine aktive Inhaltsregulierung vornehmen, selbst keine Inhalte löschen und nicht als Zensurorgan fungieren. Sie soll auch nicht als nationale „Supportabteilung" der Plattformen dienen, nicht als erste Anlaufstelle für Beschwerden fugieren und nicht bestehende plattforminterne und –externe Meldesysteme ersetzen. Sie soll primär als Berufungsorgan gegenüber Moderationsentscheidungen der Plattformen fungieren und dadurch die primäre Moderationsarbeit von Meldestellen und Plattformen für strittige Fälle ergänzen. Auch eine Streitschlichtungsfunktion könnte angemessen sein. Sie soll Berufungsfälle individuell prüfen, und Empfehlungen für den angemessenen Umgang mit den inkriminierten Inhalten durch Plattformbetreiber abgeben.

Wichtig ist, dass eine solche Ombuds- und Schlichtungsstelle keinen einseitigen Interesseneinflüssen unterworfen ist. Deshalb bietet sich ein Forum mit breiter Stakeholdereinbindung an, in dem Interessenabwägung sichergestellt ist.[121] Es ist wichtig, dass eine solche Stelle die

[121] Eine mögliche Alternative wäre ein reines unabhängiges Expertengremium.

Akzeptanz von Plattform und von staatlichen Stellen erfährt, aber gleichzeitig unabhängig von den Plattformen und staatsfern organisiert ist, da die Gefahr besteht, dass Ombudsstellen politisch besetzt und instrumentalisiert werden.

Es ist insbesondere wichtig, dass die Stelle mit den Plattformen kooperiert: Plattformen müssen bereit sein, Berufungsfälle an nationale Ombuds- und Schlichtungsstellen weiterzuleiten, die Stellen müssen von den Plattformen Informationen für Verfahren einholen können, und die Plattformen sollten Ergebnisse der Verfahren (Entscheidungen, Empfehlungen) in der Inhaltsregulierung beachten, ohne sich notwendigerweise den Beschlüssen bedingungslos zu unterwerfen. Die Letztentscheidung über Inhaltsmoderation könnte weiterhin bei den Plattformen liegen. Es ist wichtig, dass die Stelle zu Beschlüssen kommt, die aufgrund einer umsichtigen Interessenabwägung durch breite Stakeholdereinbindung Legitimität beanspruchen, egal ob im Ergebnis eine unverbindliche Empfehlung oder eine verbindliche Entscheidung steht. Ombudsstellen können so als eine von mehreren Instanzen in einem kaskadischen Beschwerdemanagement fungieren. Keinesfalls sollten durch Entscheidungen einer Ombudsstelle Rechtsansprüche von Betroffenen verloren gehen oder gefährdet werden. Der Weg zu ordentlichen Gerichten soll durch die Verfahren bei einer Ombudsstelle nicht blockiert sein.

Für die initiale Einrichtung einer solchen Ombudsstelle könnte sich als institutionelles Setting eine *Selbstregulierung in einem weiten Sinn* anbieten. Diese ist gekennzeichnet durch eine freiwillige Zusammenarbeit der verschiedenen betroffenen Stakeholder in einer gemeinsamen Struktur. Plattformen könnten sich in Form einer Selbstverpflichtung, z.B. im Rahmen eines Memorandum of Understanding, zu einer Kooperation mit den Ombudsstellen bekennen. Sie könnte als Ombuds- und Schlichtungsstelle für Inhaltsmoderation im Internet neu gegründet werden. Ihre Aufgaben könnten aber auch bei bestehenden Institutionen mit ähnlichem Profil angegliedert werden, zum Beispiel beim Österreichischen Internet Ombudsmann.

Eine Ombudsstelle sollte über eine solide finanzielle Grundlage verfügen, um ihre Aufgaben wahrnehmen zu können. Diese Finanzierung sollte aus verschiedenen Quellen erfolgen, um einseitige finanzielle Abhängigkeit zu vermeiden. Nachdem sowohl die Plattformen als auch die Öffentlichkeit von einer Ombudsstelle profitieren, könnte die Stelle durch eine Kombination aus öffentlichen Beiträgen und Beiträgen der Plattformbetreiber finanziert werden. Für eine finanzielle Beteiligung der Plattformen bestehen Anreize, weil sie durch Ombudsstellen von schwierigen inhaltsbezogenen Entscheidungen entlastet werden.[122] Auch ein Teil der für Österreich angekündigten Digitalsteuer[123] könnte für die Finanzierung einer Ombudsstelle zur Verfügung gestellt werden.

[122] Vgl. z.B. Jarvis (2019): „I suspect that the platforms would gladly pay if this got them out of the position of having to enforce vague laws with undue speed and with huge fines hanging over their heads."

[123] Das Finanzministerium rechnet mit rund 200 Mio. EUR an Einnahmen. BMF (2019), Nationale Digitalsteuer fixiert, https://www.bmf.gv.at/presse/Nationale_Digitalsteuer_fixiert_.html

Sofern ein anhaltender Bedarf an einer solchen Ombuds- und Schlichtungsstelle besteht, aber die Einrichtung oder der Betrieb im Weg der Selbstregulierung scheitern, kann ermittelt werden, ob sich eine *Ombudsstelle als Ko-Regulierung* auf Basis einer rechtlichen Grundlage anbietet. So sieht in Deutschland das Netzwerkdurchsetzungsgesetz die Möglichkeit einer regulierten Selbstregulierung für Entscheidungen über rechtswidrige Inhalte vor. In diesem Fall müsste geprüft werden, ob eine Stelle mit Ombuds- und Schlichtungsfunktionen in Österreich innerhalb einer stärker verrechtlichten Struktur verankert werden sollte, z.B. auf Basis einer einschlägigen gesetzlichen Grundlage, mit der grundsätzliche Anforderungen für Moderations- und Beschwerdeinstanzen und -verfahren festgesetzt werden. Im Rahmen einer Ko-Regulierung könnte die operative Prüfung von Inhalten und Moderationsentscheidungen ebenfalls durch eine unabhängige Prüfungsinstanz erfolgen. Die Verpflichtungen der Plattformen im Rahmen der Prüf- und Moderationsverfahren könnten auf Basis eines Verhaltenskodex geregelt werden. Im Sinne einer prozeduralen Regulierung müssten im Rahmen der Ko-Regulierung die zentralen institutionellen Komponenten wie Prüfinstanz, Verhaltenskodex und Verfahrensmodalitäten durch zuständige Aufsichtsbehörden ex ante überprüft und anerkannt und ex post evaluiert werden. Der Vorteil einer Lösung im Rahmen einer Ko-Regulierung besteht in höher Rechtssicherheit und Verbindlichkeit. Dem stehen als Nachteile u.a. langwierigere Institutionalisierungsprozesse und eine geringere Flexibilität gegenüber.

Nationale Rückkoppelung im Rahmen von Aufsicht und Kontrolle

Aufgrund der umfassenden wirtschaftlichen und politischen Macht der Plattformen, ihres starken Einflusses auf private und öffentliche Kommunikation und den aktuellen Skandalen wird verstärkt und immer häufiger die Forderung nach einer formalen Aufsicht und Kontrolle für Plattformen erhoben. Plattformen prägen mit Inhaltsregulierung durch Filterung und Kuratierung gesellschaftliche Kommunikation, die für die soziale Entwicklung und demokratische Prozesse wichtig sind. Es besteht daher weitgehende Einigkeit darüber, dass auch die Normen, Verfahren und Ergebnisse der Inhaltsregulierung der Internet-Plattformen Gegenstand regelmäßiger Überprüfungen durch externe Kontrolle sein soll. Bei dieser Aufsicht und Kontrolle geht es darum:

- die technischen und menschlichen Moderationssysteme inklusive ihrer Filter- und Kuratierungskomponenten hinsichtlich möglicher Fehlfunktionen zu überprüfen und problematische Funktionsweisen und Implikationen wie overblocking, Verzerrungen und Manipulation zu identifizieren;
- die Plattformen hinsichtlich der Einhaltung grundlegender Prinzipien und Standards im Rahmen der Inhaltsmoderation zu kontrollieren, sowie hinsichtlich ihrer Maßnahmen und Performance im Rahmen von Selbstregulierungsinitiativen;
- eine allgemeine Beobachtung von Entwicklungen in der Plattformkommunikation hinsichtlich problematischer Implikationen zu ermöglichen, die als Grundlage für weitere Reformen herangezogen werden können;

Experten-Auditing: Mix aus formaler internationaler Kontrolle und nationalen Inputs

Für die dafür nötige Aufsicht und Kontrolle ist eine funktionierende Aufgabenverteilung zwischen internationalen und nationalen Kontrollinstanzen besonders wichtig.

Internationale Aufsicht: Europäische Agency mit Kontrollkompetenzen (Auditing)

Für die Evaluierung der Performance im Rahmen bestehender internationaler Selbstregulierungsinitiativen und für die Kontrolle der technischen und menschlichen Moderationssysteme, inklusive ihrer Filter- und Kuratierungskomponenten hinsichtlich möglicher Fehlfunktionen, bietet sich eine *zentrale internationale Lösung* mit formalen Überprüfungs-/Auditfunktionen an. Dafür sollte eine internationale Instanz (Agency, Board, Panel) eigerichtet und mit offiziellen, formalen und problemadäquaten Überprüfungskompetenzen und -mandaten für Audits und Evaluierung ausgestattet werden.

Einem solchen Aufsichtsgremium könnten vor allem Expertinnen und Experten aus den Bereichen Regulierung, Wissenschaft und Zivilgesellschaft angehören. Vertrauen ist zentral, denn Plattformen müssen Zugang zu Moderationsprozessen und -daten gewähren. Ein internationales Panel könnte die Inhaltsregulierung überprüfen, ohne dass die genaue Funktionsweise der Moderation der gesamten Öffentlichkeit gegenüber offengelegt werden muss. Dadurch würden vermutlich Akzeptanz und Beteiligungsbereitschaft seitens der Plattformen steigen, und darin besteht der größte Vorteil einer zentralen, internationalen Lösung. Ein internationales und plattformübergreifendes Auditing hätte weiters den Vorteil, dass einheitliche Prüfstandards angewendet und Vergleiche ermöglicht werden (Gillespie 2018), mit denen sich best practic zeigen lässt, das in der Folge mit weiteren Plattformen geteilt werden kann.

Trotzdem ist auch im Bereich Auditing eine *nationale Rückbindung* möglich und sinnvoll. Überprüfungen sind vielfach kleinteilig, entlang spezifischer Fragestellungen und Probleme angelegt und erfordern technischen Sachverstand und inhaltliche Interessen die weltweit verstreut sind und genutzt werden sollten. Deshalb sollten auch für dezentrale Einrichtungen Optionen eröffnet werden, solche Überprüfungen durchzuführen. Dabei stellen Entscheidungen über den Umfang und die Bedingungen für den Zugang zu Systemen und Daten der Inhaltsmoderation die größte Herausforderung dar. Die Palette an Möglichkeiten reicht von offenen Schnittstellen bis zu gezielten, kontrollierten Partnerschaften. Zugangskontrolle und -beschränkungen sind notwendig, um vor Manipulation zu schützen und Datenschutz zu gewährleisten. Im Ergebnis könnte es jedoch von den Plattformen möglich gemacht werden, Audits zur Inhaltsmoderation in einem gewissen Ausmaß auch dezentral und problemgetrieben zu ermöglichen, um die Inhaltsmoderation vor dem Hintergrund *nationaler/regionaler Begebenheiten* und Bedürfnisse prüfen zu können. Dies könnte z.B. durch wissenschaftliche Expertin-

nen und Experten, interessierte Journalisten und NGOs oder im Rahmen einer nationalen Beobachtungsstelle (siehe unten) erfolgen.[124] Problematische Erkenntnisse und Kritik könnten an die Plattformen selbst, sowie an ein internationales Gremium mit formalen Prüfbefugnissen (siehe oben) weitergeleitet werden. Diese nationalen Informationen können als Input für internationale Überprüfungen fungieren. So könnte auch im Überprüfungsprozess eine nationale Rückkoppelung sichergestellt werden.

Nationale Aufsicht: Ombudsstelle mit Beobachtungs- und Evaluierungsfunktion

Eine nationale Aufsicht und Kontrolle über die internationalen Plattformen bietet Vor- und Nachteile, die für Reformüberlegungen zu beachten sind. Für nationale Aufsicht spricht u.a., dass diese leichter einzurichten ist und mit effektiven Durchsetzungsinstrumenten ausgestattet werden kann. Gegen nationale Aufsicht spricht u.a., dass zu viele, zu unterschiedliche nationale Vorgaben von Plattformen praktisch womöglich kaum bewältigt werden können.

Für Österreich scheint die Einrichtung einer eigenen neuen Aufsichtsbehörde mit weitreichenden Kompetenzen für die Kontrolle von Internet-Plattformen und sozialen Netzwerken auf Basis rechtlicher Grundlagen – wie sie beispielsweise zuletzt im Vereinigten Königreich (Bunting 2018; LSE 2018; House of Lords 2019) vorgeschlagen wurde – kurzfristig wenig realistisch.

Trotzdem soll auch im nationalen Kontext Kontrolle ermöglicht werden. Dafür bietet sich z.B. die Option, die oben vorgeschlagene unabhängige Ombuds-/Schlichtungsstelle auch mit *Beobachtungs- und Evaluierungsfunktionen* auszustatten. Eine Ombudsstelle bietet sich für zusätzliche Beobachtungsfunktionen an, weil sie aufgrund ihrer Tätigkeit mit den verschiedenen aktuellen Problemlagen vertraut ist. Darüber hinaus könnten bei einer Beobachtungsstelle Informationen zu Entwicklungen, Herausforderungen und Risiken der Plattformkommunikation und Inhaltsmoderation zusammenlaufen und von Expertinnen und Experten analysiert werden, um weiteren Regulierungsbedarf zu ermitteln. Dazu zählt beispielweise die Analyse von problematischen Phänomenen wie Desinformationskampagnen, aber auch ein Monitoring über die Anzahl, Entwicklung und Qualität von Meldungen (flagging) und Moderationsentscheidungen der Plattformen sowie die regelmäßige wissenschaftliche Evaluierung der landesspezifischen Melde- und Moderationspraxis. Eine Beobachtungsstelle sollte regelmäßig öffentliche Berichte vorlegen. Eine kompetente Beobachtungsstelle könnte sich zudem – im internationalen Verbund – an der Evaluierung der internationalen Selbstregulierungsinitiativen der Plattformen, und an den oben vorgeschlagenen Audits über Verfahren der Inhaltsmoderation beteiligen. Dadurch wäre eine nationale Rückkoppelung einer international organisierten Aufsicht sichergestellt.

[124] Vgl. dazu auch die Vorschläge zum algorithmic auditing (Diakopoulos 2015; Sandvig et al. 2014), das auch zur Überprüfung des algorithmen-basierten Flagging eingesetzt werden kann, sowie Vorschläge zur Einrichtung von Datenschnittstellen und Transparenztools (z.B. Sängerlaub 2019, 27f.)

Die Umsetzung von Beobachtungs- und Evaluierungsfunktionen erfordert Know-how[125] sowie erhebliche personelle und finanzielle Mittel. Damit stellt sich die Frage, welche Einrichtungen über die notwendigen Ressourcen verfügen. Für Österreich wurden bislang noch keine konkreten Vorschläge zur Diskussion gestellt. Einschätzungen von Expertinnen und Experten zu Folge verfügen bestehende österreichische Einrichtungen – sowohl Behörden als auch NGOs – bislang jedoch nicht über die notwendige Ressourcenausstattung, um Beobachtungs- und Evaluierungsfunktionen kurzfristig wahrnehmen zu können. Es wäre daher voraussichtlich notwendig entsprechende Kompetenz neu aufzubauen und die dafür notwendigen Ressourcen bereitzustellen. Analog zur Ausstattung einer Ombudsstelle könnten auch die Ressourcen für systematisches Monitoring und regelmäßige Evaluierung aus Beiträgen der Plattformbetreiber sowie öffentlichen Beiträgen finanziert werden, z.B. durch Anteile an Einnahmen aus der angekündigten Digitalsteuer.

Zur Ausübung ihrer Beobachtungs- und Evaluierungsfunktionen muss die Ombudsstelle u.a. relevante Informationen und Daten von den Plattformen erhalten. Dies sollte zunächst in Form von freiwilliger Kooperation und Zusammenarbeit erprobt werden. Wenn die freiwillige Kooperation scheitert, müssen Regulierungsbehörden zusätzliche Maßnahmen in Erwägung ziehen, um durch Anreize oder Druck die Kooperation der Plattformen sicher zu stellen. So könnten beispielsweise finanzielle Anreize zur Kooperation im Rahmen der Digital- oder Werbesteuer geschaffen werden. Die Werbe- oder Digitalsteuer könnte gestaffelt werden, und die Einstufung der Unternehmen kann von der Kooperation mit einer Aufsichts- und Ombudsstelle abhängig gemacht wird.[126]

Wenn freiwillige Maßnahmen und finanzielle Anreize nicht ausreichen, um die Kooperationsbereitschaft sicher zu stellen, muss geprüft werden, ob und wie eine unabhängige nationale Beobachtungsstelle mit formalen Auskunftsrechten ausgestattet werden kann. Dafür müsste geprüft werden, ob eine Beobachtungsstelle innerhalb einer stärker verrechtlichten Struktur verankert werden sollte, d.h. als *Aufsichtsstelle auf Basis einer einschlägigen gesetzlichen Grundlage* mit formaler Aufsichtsfunktion und Auskunftsrechten, sowie Transparenz- und Berichtspflichten für Meldestellen und Plattformen als Voraussetzung für die Evaluierung der Melde- und Moderationspraxis. Zudem müsste eine Aufsichtsstelle mit adäquaten Sanktionsmöglichkeiten ausgestattet sein, wenn Plattformen den Informationsaustausch verweigern.[127]

---

[125] Hilfreich für die Aufgaben einer Beobachtungsstelle sind u.a. Know-how zu rechtlichen Rahmen, zur Funktionsweise der Plattformen, sowie zur Erhebung und statistischen Auswertung der relevanten Daten.

[126] Zu geeigneten Regulierungsoptionen unter Berücksichtigung von Meinungsfreiheit und der digitalen Dominanz der Plattformen siehe Wagner (2018b, 232f.), sowie - darauf aufbauend - zur Staffelung der Werbesteuer als Anreizinstrument, Experteninterview vom 29.5.2019.

[127] Sie dazu etwa die Konstruktion im Rahmen der österreichischen Telekommunikationsregulierung mit der Telekom-Control-Kommission als unabhängiger Aufsichtsbehörde, der RTR als Geschäftsapparat und der dort angegliederten Schlichtungsstelle für Telekommunikation.

Die Wahrscheinlichkeit, dass die großen Internet-Plattformen (freiwillig) mit nationalen Einrichtungen wie Ombuds- und Beobachtungsstellen kooperieren, galt lange Zeit als sehr gering. Seitens der global agierenden Plattformen wurden möglichst zentrale, international einheitliche und wenn möglich technische Lösungen angestrebt. Mit der zunehmenden Belastung der Plattformen durch Inhaltsmoderation, und mit der Last der Verantwortung, vielfach „alleine" über Meinungsfreiheit zu entscheiden, könnten sich die Kontextbedingungen ändern und die Chancen für eine besser verteilte Verantwortung erhöhen. Solche Entwicklungen und Erkenntnisse auf Seiten der Plattformen müssen von Akteuren in nationalen Öffentlichkeiten weiter beobachtet werden, um neu entstehende Möglichkeiten für die Einrichtung nationaler Ombuds- und Beobachtungsstellen zu nutzen.

5.3 Fazit zu Reformoptionen für Österreich

Die Umsetzung eines verantwortungsorientierten Governance-Arrangements für Inhaltsregulierung auf nationaler Ebene erfordert institutionelle Reformen. Abschließend werden ausgewählte Reformoptionen identifiziert, wobei folgende Rahmenbedingungen zu beachten sind:

- *Keine „one-size-fits-all" Lösungen*: Zur Frage der Inhaltsregulierung auf Internet-Plattformen gibt es keine Universallösungen. Es besteht der Bedarf für ein Zusammenspiel zwischen den verschiedenen Governance-Akteuren und -Ebenen.
- *Kohärente Governance-Architektur:* Governance auf nationaler Ebene muss kompatibel mit bestehenden nationalen und internationalen Regelungen sein. Verantwortlichkeit auf nationaler und internationaler Ebene, sowie für private und staatliche Akteure soll klar verteilt und zugeordnet werden.
- *Enforcement Pyramide:* Harte, staatliche Regulierung soll zur Anwendung kommen, wenn freiwillige, kooperative, selbstregulative Verfahren versagen.
- *Adaptive Governance:* Die Reform des institutionellen Arrangements der Inhaltsregulierung auf Internet-Plattformen ist ein iterativer Prozess, der laufende Anpassungen ermöglichen muss.

Auf Basis der Rahmenbedingungen wurden folgende institutionelle Reformoptionen identifiziert:

Kommunikative Verantwortlichkeit: Bevollmächtigte nationale Auskunftsverantwortliche

Die Plattformen sollten Sprecher auf nationaler Ebene einsetzen, damit kommunikative Verantwortung gegenüber nationalen Öffentlichkeiten besser wahrgenommen werden kann.

Im *Melde- und Beschwerdemanagement* bietet eine nationale Rückkoppelung Vorteile für Plattformen, Nutzer und nationale Öffentlichkeiten. Dafür ist es wichtig, Melde- und Beschwerdeverfahren in Formen geteilter Verantwortung und in Phasen mehrstufiger Prozesse in „kaska-

dischen Systemen“ (z.B. Spielkamp 2018) zu etablieren. Reformen im Beschwerdemanagement würden sowohl zu einer Aufgaben- und Verantwortungsverteilung beitragen, als auch eine nationale Rückkoppelung im Beschwerdemanagement ermöglichen:

Meldewesen: Förderung und Ausbau nationaler Meldestellen
Meldestellen auf nationaler Ebene bilden wichtigen Seismographen und Knotenpunkte zwischen nationalen Öffentlichkeiten und den Plattformen. Ihr Bestand und ihre Entwicklung, die Zusammenarbeit mit den Plattformen (z.B. Trusted Flagging Programme) und ihre Beratungsangebote sollen gefördert und ihre Funktionen mit öffentlichen Mitteln sichergestellt werden.

Inhaltsmoderation: Moderatoren mit regionalem/nationalem Kontextwissen
Plattformen sollen Moderatorinnen und Moderatoren mit nationalem Kontextwissen beschäftigen, z.B. Mitarbeiterinnen und Mitarbeiter mit österreichischem Kontextwissen in Moderationsteams in Deutschland.

Berufung & Streitbeilegung: Einrichtung einer nationale Ombudsstelle
Bei etlichen Problemlagen stellt sich die Frage wie Moderationsprozesse und -entscheidungen der Plattformbetreiber beeinsprucht werden können. Dafür bietet sich die Einrichtung einer staatsfernen nationalen österreichischen Ombuds- und Schlichtungsstelle für Inhaltsmoderation im Internet an, die als Berufungsorgan gegenüber Moderationsentscheidungen der Internet-Plattformen fungiert. Die Stelle kann als Forum mit breiter Stakeholdereinbindung neu oder bei bereits bestehenden Ombuds- oder Schlichtungsstellen eingerichtet werden. Sie soll Fälle prüfen, und gegenüber Plattformen Empfehlungen aussprechen, zu denen sie aufgrund einer umsichtigen Interessenabwägung durch eine breite Stakeholdereinbindung gelangt. Als institutionelles Setting bietet sich dafür eine Selbstregulierung im weiten Sinn auf Basis von freiwilliger, kooperativer Zusammenarbeit mit den Plattformen an.

Im Bereich *Aufsicht und Kontrolle* geht es darum, die Moderationssysteme inklusive ihrer technischen Komponenten hinsichtlich möglicher Fehlfunktionen zu überprüfen; die Plattformen hinsichtlich der Einhaltung grundlegender Standards für Inhaltsmoderation und hinsichtlich ihrer Performance im Rahmen von Selbstregulierungsinitiativen zu evaluieren; und eine allgemeine Beobachtung von Entwicklungen in der Plattformkommunikation hinsichtlich problematischer Implikationen zu ermöglichen. Für die dafür nötige Aufsicht und Kontrolle ist eine funktionierende Aufgabenverteilung zwischen internationalen und nationalen Kontrollinstanzen besonders wichtig.

Internationale Aufsicht: Europäische Agency mit Kontrollkompetenzen (Auditing)
Für die Kontrolle der technischen und menschlichen Moderationssysteme, sowie für die Evaluierung der Performance im Rahmen bestehender internationaler Selbstregulierungsinitiativen bietet sich eine *zentrale internationale Instanz* (Board, Panel, Agency) mit formalen Überprüfungs-/Auditfunktionen an. Eine zentrale Lösung, z.B. auf EU-Ebene, bietet Vorteile für

Akzeptanz und Vertrauen seitens der Plattformen, Faktoren, die wichtig für die Beteiligungsbereitschaft und den Zugang zu Moderationsprozessen und -daten sind.

Nationale Kontrolle: Nationale Ombudsstelle mit Aufsichts- und Beobachtungfunktion
Auf nationaler Ebene kann eine Ombudsstelle aufgrund ihrer Problemnähe zusätzliche Aufsichts- und *Monitoringfunktionen* übernehmen, Entwicklungen und Herausforderungen der Plattformkommunikation und Inhaltsmoderation systematisch beobachten und sich im internationalen Verbund an Audits und Evaluierungen der Plattformen beteiligen. Dadurch wäre eine nationale Rückkoppelung einer international organisierten Aufsicht sichergestellt. Zur Ausübung ihrer Aufsichts- und Beobachtungsfunktionen müssen der Ombudsstelle relevante Informationen und Daten von den Plattformen in freiwilliger Kooperation zur Verfügung gestellt werden.

Von Selbst- zu Ko-Regulierung
Wenn die Einrichtung oder der Betrieb einer Ombudsstelle mit Aufsichts- und Beobachtungsfunktionen im Weg der freiwilligen Kooperation scheitert, müssen Regulierungsbehörden zusätzliche Maßnahmen in Erwägung ziehen, um die Kooperation der Plattformen sicher zu stellen. So könnten beispielsweise Anreize zur Kooperation geschaffen werden, indem die Werbe- oder Digitalsteuer gestaffelt, und die Höhe von der Kooperation mit einer Aufsichts- und Ombudsstelle abhängig gemacht wird. Wenn die freiwillige Kooperation scheitert sollte zudem ermittelt werden, ob eine entsprechende Aufsichtsstelle innerhalb einer stärker verrechtlichten Struktur verankert werden sollte, z.B. als Ko-Regulierung auf Basis einer einschlägigen gesetzlichen Grundlage, die adäquate Sanktionsmöglichkeiten vorsieht.

Literaturverzeichnis

Article 19 (2018) *Regulating social media: we need a new model that protects free expression*. https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/

Article 19 (2018). *Self-regulation and 'hate speech' on social media platforms*. https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-'hate-speech'-on-social-media-platforms_March2018.pdf

Ash, T.G., Gorwa, R., & Metaxa, D. (2019). Glasnost! Nine ways Facebook can make itself a better forum for free speech and democracy. An Oxford-Stanford Report. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Garton_Ash_et_al_Facebook_report_FINAL_0.pdf

Ayres J., & Braithwaite , J. (1992). Responsive regulation. Transcending the Deregulation Debate. Oxford.

Berka, W., & Trappel, J. (2019). Internetfreiheit in Österreich. Eine Bestandsaufnahme auf der Grundlage der Empfehlung CM/REC(2016)5 des Ministerkomitees des Europarats an die Mitgliedstaaten zur Internetfreiheit. Wien: Manz.

BMF (2019), Nationale Digitalsteuer fixiert. Pressemitteilung vom 11.01.2019, https://www.bmf.gv.at/presse/Nationale_Digitalsteuer_fixiert_.html

Brodnig, I. (2016). Wie Algorithmen Populisten helfen – das aber nicht so sein muss. Brodnigs Blog, 13.11.2016. https://www.brodnig.org/2016/11/13/wie-algorithmen-populisten-helfen-das-aber-nicht-so-sein-muss/

Brodnig, I. (2019). Facebook regulieren: Wie viel Aufsicht darf es sein? derStandard.at, 4.7.2019. https://www.derstandard.at/story/2000105879585/facebook-regulieren-wie-viel-aufsicht-darf-es-sein

Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. European Law Journal, 13(4), 447-468.

Bovens, M. (2010). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. West European Politics, 33 (5), 946-967.

Bowcott, O. & Gibbs, S. (2017). UK considers internet ombudsman to deal with abuse complaints. The Guardian, 22.8.2017. https://www.theguardian.com/technology/2017/aug/22/uk-considers-internet-ombudsman-to-deal-with-abuse-complaints

Bradley Ch., & Wingfield, R. (2018). A Rights-Respecting Model of Online Content Regulation by Platforms. Global Partners Digital. https://www.business-humanrights.org/sites/default/files/documents/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf

Bundesjustizamt (2018). Netzwerkdurchsetzungsgesetz - Häufig gestellte Fragen. https://www.bundesjustizamt.de/DE/Themen/Buergerdienste/NetzDG/Fragen/FAQ_node.html#faq10018916

Bundesrat Schweiz (2009): Jugend und Gewalt. Wirksame Prävention in den Bereichen Familie, Schule, Sozialraum und Medien. http://www.news.- admin.ch/NSBSubscriber/message/attachments/15741.pdf.

Bunting, M. (2018). *Keeping Consumers Safe Online: Legislating for platform accountability for online content*. Report, Communication Chambers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3274556

Bunting, M. (2018) From editorial obligation to procedural accountability: policy approaches to online content in the era of information intermediaries, Journal of Cyber Policy, 3 (2), 165-186, DOI: 10.1080/23738871.2018.1519030

Burgard-Arp, N. (2018). Magersucht: Die Freundin, die sie hungern ließ. Zeit.de, 30.9.2018. https://www.zeit.de/campus/2018-09/magersucht-bulimie-pro-ana-zugehoerigkeit-social-media/komplettansicht

Caplan, R. (2018). Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches. Report, Data & Society,

Council of Europe Committee of experts on internet intermediaries (MSI-NET) (2016). *Draft Recommendation CM/Rec(2017)xxx of the Committee of Ministers to member states on the roles and responsibilities of internet intermediaries*. https://rm.coe.int/draft-recommendation-on-internet-intermediaries-7th-revised-version-/1680770c37

Council of Europe Committee of experts on internet intermediaries (MSI-NET) (2018). *Algorithms and Human Rights - Study on the human rights dimension of automated data processing techniques and possible regulatory implications*. https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5

Dachwitz, I. (2019). Warum Künstliche Intelligenz Facebooks Moderationsprobleme nicht lösen kann, ohne neue zu schaffen. Netzpolitik.org, 04.05.2019. https://netzpolitik.org/2019/warum-kuenstliche-intelligenz-facebooks-moderationsprobleme-nicht-loesen-kann-ohne-neue-zu-schaffen/

Der Standard (2018), Regierung organisiert Gipfel gegen Hass. derStandard.at, 28.10.2018, https://derstandard.at/2000090200113/Regierung-organisiert-Gipfel-gegen-Hass-im-Netz

Diakopoulos, N. (2015). Algorithmic Accountability. Journalistic investigation of computational power structures. Digital Journalism, 3 (3), 398-415.

Doteveryone (2018). Regulating for Responsible Technology. https://doteveryone.org.uk/wp-content/uploads/2018/10/Doteveryone-Regulating-for-Responsible-Tech-Report.pdf

DW – Deutsche Welle (2013). Selbstregulierung von Online-Netzwerken gescheitert. DW, 06.05.2013, https://www.dw.com/de/selbstregulierung-von-online-netzwerken-gescheitert/a-16795312

Dzakula, J. (2018), Anticipatory regulation: a way forward for platform governance? LSE Blog, 4.10.2018. http://blogs.lse.ac.uk/mediapolicyproject/2018/10/04/anticipatory-regulation-a-way-forward-for-platform-governance/

Eberwein, T., Fengler, S., & Karmasin, M. (Eds.). (2018). The European handbook of media accountability. London, New York: Routledge.

Engeler, M (2018). Meinungsfreiheit: Warum Facebook (zu Recht) nicht an Grundrechte gebunden ist. Netzpolitik.org, 18.9.2018, https://netzpolitik.org/2018/meinungsfreiheit-warum-facebook-zu-recht-nicht-an-grundrechte-gebunden-ist/

Europäische Kommission (2015). Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Eine Strategie für einen digitalen Binnenmarkt für Europa. KOM (2015) 192 final, vom 6.5.2015.

Europäische Kommission (2016a). Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Online-Plattformen im digitalen Binnenmarkt. Chancen und Herausforderungen für Europa. KOM (2016) 288 final, vom 25.5.2016

Europäische Kommission (2016b). Kommission aktualisiert EU-Bestimmungen für audiovisuellen Bereich und stellt gezielten Ansatz für Online-Plattformen vor. Pressemitteilung, 25. Mai 2016. http://europa.eu/rapid/press-release_IP-16-1873_de.htm

Europäische Kommission (2016c). Europäische Kommission und IT-Unternehmen geben Verhaltenskodex zur Bekämpfung illegaler Hassrede im Internet bekannt. Pressemitteilung, 31. Mai 2016. http://europa.eu/rapid/press-release_IP-16-1937_de.htm

Europäische Kommission (2016d). Erste Fortschritte im Kampf gegen Online-Hetze. Pressemitteilung, 06. Dezember 2016. https://ec.europa.eu/germany/news/erste-fortschritte-im-kampf-gegen-online-hetze_de

Europäische Kommission (2016e). Kommission aktualisiert EU-Bestimmungen für audiovisuellen Bereich und nimmt gezielten Ansatz für Online-Plattformen an. Factsheet, 25.05.2016. http://europa.eu/rapid/press-release_MEMO-16-1895_de.htm

Europäische Kommission (2017). Bekämpfung von Hetze im Internet– Fortschritte bei der gemeinsamen Initiative der Kommission mit Social-Media-Plattformen und der Zivilgesellschaft. Pressemitteilung, 1. Juni 2017. http://europa.eu/rapid/press-release_IP-17-1471_de.htm

Europäische Kommission (2018a). Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Bekämpfung von Desinformation im Internet: Ein europäisches Konzept. KOM (2018) 236 vom 26.4.2018.

Europäische Kommission (2018b). Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Freie und faire Europawahlen gewährleisten. KOM (2018) 637 vom 12.9.2018.

Europäische Kommission (2018c). *Statement by Commissioner Gabriel on the Code of Practice on Online Disinformation*. http://europa.eu/rapid/press-release_STATEMENT-18-5914_en.htm

Europäische Kommission (2018d). Digital Single Market: updated audiovisual rules. Memo, 7. Juni 2018 http://europa.eu/rapid/press-release_MEMO-18-4093_en.htm

Europäische Kommission (2018e), Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Verhinderung der Verbreitung terroristischer Online-Inhalte. KOM (2018) 640 vom 12.9.2018, https://ec.europa.eu/transparency/regdoc/rep/1/2018/DE/COM-2018-640-F1-DE-MAIN-PART-1.PDF

EU Code of Practice on Disinformation. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454

Facebook (2018). Begünstigung und Propagierung von Verbrechen, Gemeinschaftsstandards. https://www.facebook.com/communitystandards/promoting_publicizing_crime

Facebook (2018). Facebook Transparency Report. https://transparency.facebook.com/community-standards-enforcement

Facebook (2018). Gemeinschaftsstandards - Nacktdarstellung von Kindern und deren sexuelle Ausbeutung. https://www.facebook.com/communitystandards/child_nudity_sexual_exploitation

Facebook (2019a). An Oversight Board for Content Decisions (Draft Charter), Facebook, 28 January 2019, https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-2.pdf

Facebook (2019b). Global Feedback & Input on the Facebook Oversight Board for Content Decisions, Facebook, 27. Juni 2019. https://fbnewsroomus.files.wordpress.com/2019/06/oversight-board-consultation-report-2.pdf

Facebook (o.J.). Community Standards Enforcement Report https://transparency.facebook.com/community-standards-enforcement

Facebook (o.J). Community Standards. https://de-de.facebook.com/communitystandards/introduction

Fanta, A. (2018). Schlag gegen die Netzfreiheit: EU-Abgeordnete treffen Vorentscheid für Uploadfilter und Leistungsschutzrecht. Netzpolitik.org, 20.6.2018. https://netzpolitik.org/2018/schlag-gegen-die-netzfreiheit-eu-abgeordnete-treffen-vorentscheid-fuer-uploadfilter-und-leistungsschutzrecht/ ;

Fengler, S. Eberwein, T. & Leppik-Bork, T. (2011). Mapping Media Accountability – in Europe and Beyond. Köln: Halem.

Fiedler, K. (2019). Plattformen: Die Zukunft von Notice & Takedown in Europa. Netzpolitik.org, 07.03.2019. https://netzpolitik.org/2019/plattformen-die-zukunft-von-notice-takedown-in-europa/

Financial Times (2018). Ofcom outlines case for regulating social media networks. https://www.ft.com/content/a16935a4-bb39-11e8-94b2-17176fbf93f5

Fisher, E. C. (2014). The European Union in the Age of Accountability. Oxford Journal of Legal Studies, 24 (3), 495-515.

Forgó, N., Johnson, E., & Polychronopouls, K. (2019). Legal Review of the Austrian draft Law on Diligence and Responsibility Online. Report commissioned by the Office of the OSCE Representative on Freedom of the Media. Vienna: OSCE.

Geiger, S. (2016). Bot-Based collective Blocklists on Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. Information, Communication, and Society 19(6), 787-803.

Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG) (2017).

Gillespie, T. (2018). Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape social media. New Heaven / London: Yale University Press.

Gillespie, T. (2018). Improving Moderation. http://www.custodiansoftheinternet.org/application/files/5815/3185/4676/Gillespie_-_improving_moderation_-_extended_CH8.pdf

Gillespie, T. (2018). Platforms are not intermediaries. *Georgetown Law Technology Review*, *2*(2).

Gollatz, K., Beer, F., & Katzenbach, C. (2018). *The Turn to Artificial Intelligence in Governing Communication Online: HIIG Workshop Report*. https://www.hiig.de/wp-content/uploads/2018/09/Workshop-Report-2018-Turn-to-AI.pdf

Google (2018). YouTube-Community-Richtlinien und ihre Anwendung. Wichtige Richtlinien: Gewalttätiger Extremismus. https://transparencyreport.google.com/youtube-policy/featured-policies/violent-extremism

Gostomzyk, T. (2019). Cyber Courts – Der Schutz der Persönlichkeit unter den Bedingungen privat- öffentlicher Netzkommunikation. Vortrag, Österreichische Akademie der Wissenschaften, Wien: 11.6.2019.

Helberger, N. (2018). The political power of online platforms – too big to regulate? Presentation, University of Zurich, 13.11.2018.

Helberger, N., Pierson, J., & Poell, T. (2017). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, *34*(1), 1-14. doi: 10.1080/01972243.2017.1391913

Heldt, A. P. (2019). Upload-Filters. Bypassing Classical Concepts of Censorship? Journal of Intellectual Property, Information Technology and Electronic Commerce Law (JIPITEC), 10 (2019), 57-65.

HM Government (UK) (2019). Online Harms White Paper. https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper

House of Commons (UK), Digital, Cultural, Media and Sport Committee (2019). Disinformation and "fake news": Final Report. HC 1791, 18. February 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf

House of Commons (UK), Digital, Cultural, Media and Sport Committee (2018). Algorithms in decision-making. Final Report. HC 351, 23. May 2018. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

House of Lords: Select Committee on Communications (2017). Growing up with the internet. HL Paper 130, 21.3.2017. https://publications.parliament.uk/pa/ld201617/ldselect/ldcomuni/130/130.pdf

House of Lords (UK), Select Committee on Communications (2019). Regulating in a digital world. HL Paper 299, 9. March 2019. https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf

Hurtz, S. (2017), Bald kann jeder Facebook-Nutzer entscheiden, wie viel nackte Haut er sieht. Süddeutsche.de, 17.2.2017. https://www.sueddeutsche.de/digital/mark-zuckerberg-bald-kann-jeder-facebook-nutzer-entscheiden-wie-viel-nackte-haut-er-sieht-1.3384650

ISA – Institut für Strategieanalysen (2017). Digitalmonitor, Welle 1 Juni 2017. https://www.digitales.oesterreich.gv.at/documents/22124/30428/digitalmonitor+welle+1.pdf/b1d857f0-970f-4ea1-967f-f46503da8d54

Jarren, O. (2018). Kommunikationspolitik für die Kommunikationsgesellschaft. Verantwortungskultur durch Regulierung. Aus Politik und Zeitgeschichte 40-41 (2018). http://www.bpb.de/apuz/276557/kommunikationspolitik-fuer-die-kommunikationsgesellschaft?p=all

Jarvis, J. (2019). Proposals for Reasonable Technology Regulation and an Internet Court. Medium.com, 1. April 2019, https://medium.com/whither-news/proposals-for-reasonable-technology-regulation-and-an-internet-court-58ac99bec420

Jugendschutz.net (2016). Löschung rechtswidriger Hassbeiträge bei YouTube. Test zeigt erheblichen Nachholbedarf beim Umgang mit User-Beschwerden. https://www.fair-im-netz.de/WebS/NHS/SharedDocs/Downloads/DE/09262016_Testergebnisse_YouTube.pdf?__blob=publicationFile&v=3

Jugendschutz.net (2018). Jugendliche sicher in Social Media? Bericht. http://www.jugendschutz.net/fileadmin/download/pdf/Bericht_Social_Media.pdf

Just, N., & Latzer, M. (2017). Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet. Media, Culture & Society, 39 (2), 238–258.

Just, N., Latzer, M., & Saurwein, F. (2007). Communications Governance: Entscheidungshilfe für die Wahl des Regulierungsarrangements am Beispiel Spam. In P. Donges (Ed.), Von der Medienpolitik zur Media Governance? (pp. 103–126). Köln: Herbert von Halem Verlag.

Kadri, T.E., & Klonick, K. (2019). Facebook v. Sullivan: Building Constitutional Law for Online Speech (January 1, 2019). Southern California Law Review, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3332530 or http://dx.doi.org/10.2139/ssrn.3332530

Karmasin, M., & Litschka, M. (2017). CSR as an economic, ethical, and communicative concept. In S. Diehl, M. Karmasin, B. Mueller, R. Terlutter, & F. Weder (Eds.), Handbook of integrated CSR communication (pp. 37-51). Heidelberg/New York/Dordrecht/London: Springer.

Keller, D. (2015). Empirical evidence of „over removal" by Internet companies under intermediary liability laws. The Center for Internet and Society at Stanford Law School (Blog), October 12, 2015, http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws

Keller, D. (2018). Internet Platforms: Observations on Speech, Danger, and Money. Aegis Series Paper No. 1807.

Kelller, D. (2019). Who Do You Sue? State and Platform Hybrid Power Over Online Speech. Aegis Series Paper No. 1902. https://www.lawfareblog.com/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech

Klonick, K. (2017). The New Governors: The People, Rules, and Processes Governing Online Speech (March 20, 2017). 131 Harv. L. Rev. 1598. https://ssrn.com/abstract=2937985

Krempl, S. (2018). Erste Löschberichte bestätigen Gefahr von Overblocking. Golem.de, 31.7.2018. https://www.golem.de/news/netzdg-kritiker-erste-loeschberichte-bestaetigen-gefahr-von-overblocking-1807-135758.html

Ladeur, K.-H. (2014). Cyber Courts. Private Rechtsprechung in den neuen Medien. In: P. Felixberger, & A. Nassehi (Hrsg.), Kursbuch 177 – „Privat 2.0". Murmann Publishers.

Ladeur, K.-H. (2017). Die Medienfreiheit im Netzwerk der Netzwerke. Neue Institutionen für neue Medien. In: S. Knaller, & D. Pichler (Eds.), Recht im medialen Feld. Aktuelle und historische Konstellationen (pp. 4-25). PhiN, Beiheft 12/2017.

Latzer, M., Saurwein, F., Dörr, K., Just, N., & Wallace, J. (2015). Evaluation der Selbstregulierungsmassnahmen zum Jugendmedienschutz der Branchen Film, Computerspiele, Telekommunikation und Internet. Forschungsbericht im Auftrag des Bundesamtes für Sozialversicherungen (BSV). Zürich: IPMZ.

Latzer, M., Just, N., & Saurwein, F. (2012). Self- and Co-Regulation: Evidence, legitimacy and governance choice. In: M. E. Price, S. G. Verhulst, & L. Morgan (Eds.), The Routledge Handbook of Media Law (pp.373-397). Abingdon/New York: Routledge.

Latzer, M., Just, N., Hollnbuchner, K. & Saurwein, F. (2016). The Economics of Algorithmic Selection. In: J. M. Bauer, & M. Latzer (Eds.): Handbook on the Economics of the Internet (pp. 395-425). Cheltenham, UK and Northampton, MA: Edward Elgar.

Latzer, M., Just, N., Saurwein, F. & Slominski P. (2002). Selbst- und Ko-Regulierung im Mediamatiksektor. Alternative Regulierungsformen zwischen Staat und Markt. Wiesbaden: Westdeutscher Verlag.

Latzer, M., Just, N., Saurwein, F. & Slominski P. (2003). Regulation Remixed: Institutional Change through Self- and Co-Regulation in the Mediamatics Sector. Communications & Strategies, 50 (2) 2003, 127–157.

Latzer, M., Price, M.E, Saurwein, F., & Verhulst, S. G. (2007). Comparative Analysis of Self- and Co-Regulation in Communication Markets. Research report commissioned by the Office of Communications (Ofcom, UK). Vienna: ITA.

Latzer, M., Saurwein, F., & Just, N. (2019). Governance-Choice Method: In Search of the Appropriate Level of State Intervention. In: H. Van den Bulck & M. Puppis (Eds.), *Palgrave Handbook of Methods for Media Policy Research*. Palgrave.

LSE Commission on Truth, Trust and Technology (2018). *Tackling the Information Crisis: A Policy Framework for Media System Resilience*. http://www.lse.ac.uk/media-and-communications/assets/documents/research/T3-Report-Tackling-the-Information-Crisis.pdf

McKirdy, E, (2018). Facebook: We didn't do enough to prevent Myanmar violence. CNN Business, 6.11.2018, https://edition.cnn.com/2018/11/06/tech/facebook-myanmar-report/index.html

McSherry, C., York, J.C., & Cohn, C. (2018). Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas. Electronic Frontier Foundation, 30.1.2018. https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some

Meineck, S. (2018), Facebook erklärt jetzt so genau wie nie, was verboten ist – und wirft noch mehr Fragen auf. Motherboard, 24.4.2018. https://motherboard.vice.com/de/article/kzxjqa/facebook-erklart-jetzt-so-genau-wie-nie-was-verboten-ist-und-wirft-noch-mehr-fragen-auf

Moor M., & Tambini, D. (2018). Digital Dominance: the power of Google, Amazon, Facebook and Apple. Oxford University Press.

Newton, C. (2019). The Trauma Floor. The secret lives of Facebook moderators in America. The Verge, 25.2.2019. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

Mozilla Foundation (2019). Open letter to Facebook: Do your part against disinformation. https://blog.mozilla.org/blog/2019/02/11/open-letter-facebook-do-your-part-against-disinformation/

Newman, Abraham L., & Bach, D. (2004). Self-Regulatory Trajectories in the Shadow of Public Power, Governance 17: 387–413.

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences*, *376*(2133), 20180089. doi: 10.1098/rsta.2018.0089

Neuwal.com (2018). Interview Blümel, ZIB 2, 14. November 2018 (Transkript). https://neuwal.com/transkript/20181113-armin-wolf-gernot-bluemel-zib2.php

OFCOM (2018). *Addressing harmful online content: A perspective from broadcasting and on-demand standards regulation*. https://www.ofcom.org.uk/phones-telecoms-and-internet/information-for-industry/internet-policy/addressing-harmful-online-content

ORF.at (2018). EU wappnet sich gegen „Fake News". ORF.at, 6.12.2018. https://orf.at/stories/3102259/

ORF.at (2018). Gehversuche mit dem Anti-Hass-Gesetz. ORF.at, 20.12.2018. https://oe1.orf.at/artikel/653750

ORF.at (2018). Schutz oder Zensur?: Mit Gesetzen gegen „Fake News". ORF.at, 21.11.2018. https://orf.at/stories/3101553/

ORF.at (2019). Gesetz gegen Hass im Netz: Blümel sondiert in Paris. ORF.at, 22. März 2019, https://orf.at/stories/3116114/

Perrin, W., & Woods, L. (2018). Harm reduction in social media – what can we learn from other models of regulation? https://www.carnegieuktrust.org.uk/blog/harm-reduction-social-media-can-learn-models-regulation/

Perrin, W., & Woods, L. (2018). Reducing harm in social media through a duty of care. https://www.carnegieuktrust.org.uk/blog/reducing-harm-social-media-duty-care/

Pichler, G. (2018). "Komm in die Gruppe": Wenn auf Whatsapp der Reichtum winkt. derStandard.at, 15.2.2018. https://derstandard.at/2000072859462/Komm-in-die-Gruppe-Wenn-auf-Whatsapp-der-Reichtum-winkt

Rahmenbeschluss 2008/931/JI zur strafrechtlichen Bekämpfung bestimmter Formen und Ausdrucksweisen von Rassismus und Fremdenfeindlichkeit ABl L 328, 55.

Rat der Europäischen Union (2018), Interinstitutional File 2018/0331(COD) http://data.consilium.europa.eu/doc/document/ST-15336-2018-INIT/en/pdf

Rechteinfach.at (2015). Providerhaftung. http://www.rechteinfach.at/rechtslexikon/providerhaftung-195.html

Rudl, T. (2018). EU-Kommission will von Plattformen „freiwillige" und weitreichende Internetzensur. Netzpolitik.org, 2.3.2018. https://netzpolitik.org/2018/eu-kommission-will-von-plattformen-freiwillige-und-weitreichende-internetzensur/

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). An Algorithm Audit. In: S. P. Gangadharan (Ed.), Data and Discrimination: Collected Essays. Washington, DC: New America Foundation, 6-10.

Sängerlaub, A. (2019). Der blinde Fleck digitaler Öffentlichkeiten. Warum sich Desinformationskampagnen in sozialen Netzwerken kaum systematisch analysieren lassen. Berlin: Stiftung Neue Verantwortung.

Santa Clara Principles on Transparency and Accountability in Content Moderation (2018). https://santaclaraprinciples.org/

Saurwein, F. (2011). Regulatory Choice for Alternative Modes of Regulation: How Context Matters. Law & Policy, 33 (3) 2011, 334–366.

Saurwein, F. (2019). Automatisierung, Algorithmen, Accountability: Eine Governance Perspektive. In: M. Rath, F. Krotz, & M. Karmasin (Hrsg.), Maschinenethik. Normative Grenzen autonomer Systeme. Springer.

Saurwein, F., & Latzer, M. (2010). Regulatory Choice in Communications: The Case of Content-Rating Schemes in the Audiovisual Industry. Journal of Broadcasting & Electronic Media, 54 (3) 2010, 463–484.

Saurwein, F., Just, N., & Latzer, M. (2015). Governance of Algorithms: Options and Limitations. Info - The journal of policy, regulation and strategy for telecommunications, information and media, 17 (6) 2015, 35–49.

Saferinternet.at (2017): Aktiv gegen Hasspostings (Leitfaden), https://www.saferinternet.at/fileadmin/categorized/Materialien/Aktiv_gegen_Hasspostings_Leitfaden.pdf

Scott, C. (2000). Accountability in the Regulatory State. In: C. Harvey, J. Morison, & J. Shaw, (Eds.), Voices, Spaces and Processes in Constitutionalism (pp. 38-60). London: Blackwell.

Sombetzki, J. (2016). Verantwortung und Roboterethik - ein kleiner Überblick. Humboldt Forum Recht (HFR), 3/2016, 10-30.

Sounding Board of the Multistakeholder Forum on Disinformation Online. (2018). *The Sounding Board's Unanimous Final Opinion on the So-Called Code of Practice*. https://www.euractiv.com/wp-content/uploads/sites/2/2018/10/3OpinionoftheSoundingboard-1.pdf

Spielkamp, M. (2018). Why Zuckerberg's "Independent Governance and Oversight" board is not gonna fly (but some of his other ideas are at least worth discussing). AlgorithmWatch, 17.11.2018. https://algorithmwatch.org/en/why-facebooks-independent-governance-and-oversight-board-is-not-gonna-fly-but-some-of-his-other-ideas-are-at-least-worth-discussing/

Stadler, T. (2018). Die Debatte um das NetzDG ist unsachlich. Internet-Law.de (Blog), http://www.internet-law.de/2018/01/die-debatte-um-das-netzdg-ist-unsachlich.html

Stadler, T. (2016). Nach welchen Kriterien löscht Facebook? Internet-Law.de (Blog), 25.1.2016. http://www.internet-law.de/2016/11/nach-welchen-kriterien-loescht-facebook.html

Tambini, D. (2018). Long Read: What would a social media regulator actually do? Inforrm's Blog, The International Forum for Responsible Media Blog, 22.09.2018. https://inforrm.org/2018/09/22/long-read-what-would-a-social-media-regulator-actually-do-damian-tambini/

Thompson, N. (2017). Instagram Unleashes an AI System to Blast Away Nasty Comments. Wired, 29.6.2017. https://www.wired.com/story/instagram-launches-ai-system-to-blast-nasty-comments/

Tschohl, Ch. (2016). Studie zum Konzept einer zentralen „Clearingstelle" zur inhaltlichen Beurteilung von Netzsperren im Zusammenhang mit Verletzungen des Urheberrechts. Studie im Auftrag der Internet Service Provider Austria (ISPA). Wien: Research Institute.

Twitter (2018). Twitter Rules enforcement. https://transparency.twitter.com/en/twitter-rules-enforcement.html

Twitter (o.J.). Die Twitter Regeln. https://help.twitter.com/de/rules-and-policies/twitter-rules

Van Zuylen-Wood, S. (2019). "Man are scum": Inside Facebook's war on hate speech. Vanity fair, 26. February 2019, https://www.vanityfair.com/news/2019/02/men-are-scum-inside-facebook-war-on-hate-speech

Verhaltenskodex für die Bekämpfung illegaler Hassreden im Internet. http://ec.europa.eu/newsroom/document.cfm?doc_id=42875

Wagner, B. (2018a). Ethics as an Escape from Regulation: From ethics-washing to ethics shopping? In: M. Hildebrandt (Ed.), Being Profiling. Cogitas ergo sum. Amsterdam University Press.

Wagner, B. (2018b). Free Expression? Dominant information intermediaries as arbiters of internet speech. In M. Moore, & D. Tambini (Eds.), Digital Dominance. The power of Google, Amazon, Facebook and Apple (pp. 219-240). New York, NY: Oxford University Press.

Wagner, B. (2019). Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems. Policy & Internet, online first, 24 January 2019, https://doi.org/10.1002/poi3.198

Wallner, A. (2018). Hass statt Honig für Wiener Neujahrsbaby Asel. Die Presse, 3.1.2018. https://diepresse.com/home/meinung/kommentare/5347469/Hass-statt-Honig-fuer-Wiener-Neujahrsbaby-Asel

Wannenmacher, T. (2018). Falsche "ZIB" Werbung lockt Nutzer in eine Falle! Mimikama, 14.12.2018. https://www.mimikama.at/allgemein/fake-zib/

Warner, M. (2018). *Potential Policy Proposals for Regulation of Social Media and Technology Firms*. https://regmedia.co.uk/2018/07/30/warner_social_media_proposal.pdf

Waterson, J. (2018). Ofcom to push for regulation of social networks, The Guardian, 17.9 2018. https://www.theguardian.com/media/2018/sep/17/ofcom-to-push-for-regulation-of-social-media-platforms

Weder, F., & Karmasin, M. (2011). Corporate communicative responsibility: Kommunikation als Ziel und Mittel unternehmerischer Verantwortungswahrnehmung. Studienergebnisse aus Österreich. Zeitschrift für Wirtschafts- und Unternehmensethik, 12 (3), 410-428.

Windhager, M., & Forgó, N. (2017), Hasspostings , Bedarf es einer gesetzlichen Verschärfung? AnwBl 10/2017, 592.

York, J. & Wagenknecht, A. (2016). Nudes and N00dz. re:publica 2016, https://re-publica.com/de/session/nudes-and-n00dz

YouTube (2019). Continuing our work to improve recommendations on YouTube. YouTube, Official Blog, 25.1.2019. https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html

Zankl, W. (2016). ECG - E-Commerce-Gesetz. Kommentar. 2. Auflage. Wien: Verlag Österreich.

ZARA – Zivilcourage und Anti-Rassismus-Arbeit (2018). Rassismus Report 2017. Einzelfall-Bericht über rassistische Übergriffe und Strukturen in Österreich. Wien: ZARA.

ZARA – Zivilcourage und Anti-Rassismus-Arbeit (2019). Rassismus Report 2018. Einzelfall-Bericht über rassistische Übergriffe und Strukturen in Österreich. Wien: ZARA.

Zuckerberg, M. (2017). Building Global Community. Facebook, 16.2.2017. https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10103508221158471/

Zuckerberg, M. (2018). A Blueprint for Content Governance and Enforcement. https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/

Zuckerberg (2019), A Privacy-Focused Vision for Social Networking, https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/

WWW.OEAW.AC.AT/CMC