

Paradigmen der Plattformpolitik im Umgang mit Hass und Extremismus: eine chronologische Einordnung

Maik Fielitz & Marcel Jaspert (Institut für Demokratie und Zivilgesellschaft)

Dieser Beitrag gibt einen Überblick über den Umgang von Tech-Unternehmen mit der Verbreitung von Hass und Extremismus auf ihren digitalen Plattformen. Digitale Plattformen werden dabei als politische Akteure greifbar gemacht, da sie vielschichtige soziale Beziehungen strukturieren, kuratieren und weitreichende politische Entscheidungen treffen können. Plattformen sind Austragungsorte gesellschaftlicher Konflikte, auf denen radikalisierte Inhalte leicht Verbreitung finden, sodass sie auch Tummelplätze für Extremist*innen geworden sind. Da liberale Demokratien einen Umgang mit diesem Hass und Extremismus fordern, sind die Plattformen gefordert, mit Maßnahmen dagegen zu reagieren. In diesem Beitrag werden auf Basis von 403 Blogartikeln der Unternehmensblogs von YouTube, Twitter und Facebook sowie einer qualitativen Inhaltsanalyse Determinanten herausgearbeitet, auf die die Plattformen in ihren politischen Entscheidungen bisher reagiert haben. Die Determinanten Regulierung, ökonomische Interessen, technische Möglichkeiten, Erwartungen und Bedürfnisse der Nutzer*innen und gesellschaftliche Werte und Normen werden fundiert dargestellt. Darüber hinaus werden verschiedene Paradigmen der Plattformpolitik in vier Perioden herausgearbeitet und der dynamische Prozess sichtbar gemacht.

Empfohlene Zitierung:

Fielitz Maik/Jaspert Marcel (2023). Paradigmen der Plattformpolitik im Umgang mit Hass und Extremismus: eine chronologische Einordnung. In: Institut für Demokratie und Zivilgesellschaft (Hg.). Wissen schafft Demokratie. Schwerpunkt Netzkulturen und Plattformpolitiken, Band 14. Jena, 106–123.

Schlagwörter:

Plattformpolitik, Hass, Extremismus, Regulierung



**„OB GEWOLLT ODER NICHT,
BETREIBEN DIE KONZERNE
PLATTFORMPOLITIK – IN
FORM DER ORGANISIERUNG
IHRER MACHTBASIS UND IN
FORM DER EINFLUSSNAHME
AUF POLITISCHE SYSTEME
UND DIE TRANSNATIONALE
NUTZERGEMEINSCHAFT.
DIES HAT AUSWIRKUNGEN
DARAUF, WIE SICH AKTEURE
UND IDEOLOGIEN VERBREITEN
UND EINE BREITE MASSE
ERREICHEN.“**

Maik Fielitz & Marcel Jaspert

Einleitung

Im Januar 2021 trafen die CEOs der wichtigsten globalen Plattformunternehmen eine weitreichende Entscheidung: Sie verbannten den abgewählten US-Präsidenten Donald J. Trump von ihren Plattformen, die er zuvor Jahre lang nutzte, um politische Propaganda zu verbreiten. Nach dem Sturm auf das Kapitol im Januar 2021, der fünf Menschen das Leben kostete und einen beispiellosen Angriff auf die Institutionen der liberalen Demokratie in den USA darstellte, befanden die Verantwortlichen von Facebook, Twitter und YouTube, dass Trump die Gewalt billigte und verbannten ihn von ihren Diensten – vorerst. Denn nur zwei Jahre später wurde Trump seine Rückkehr in den digitalen Mainstream ermöglicht, nachdem das Oversight Board von Meta (früher Facebook) entschied, dass von Trump keine Gefahr mehr ausgehe. Auch andere Plattformen schlossen sich dieser Lesart an (Jaspert 2023).

Der Fall Trump stellt zwei Zäsuren in der globalen Plattformpolitik dar, die nur im Kontext sich wandelnder politischer Kräfteverhältnisse zu verstehen sind: Während es nach der Abwahl als Präsident politisch opportun war, Trump zu isolieren, da er kein politisches Amt mehr innehatte, führte die Aussicht, er könne erneut Präsident werden zur Neubewertung der Situation. Denn nicht nur wurden Trumps Accounts reaktiviert, sondern auch die Richtlinien geändert. Danach ist die Mär von der geklauten Wahl nun wieder eine legitime Meinung auf der Plattform und keine gefährliche Verschwörungstheorie, die politische Gewalt nach sich ziehen könnte (Bond 2023). Hinzu kam ein Paradigmenwechsel an der Spitze des Twitter-Konzerns, der freie Rede nun so weitreichend definiert, dass extremistische Akteure wie Andrew Anglin, Martin Sellner und Kanye West wieder ein Massenpublikum adressieren können (Stokel-Walker 2022).

Die Regeln des digitalen Austauschs ändern sich stetig und hängen von Präferenzen der Plattformen ab. Mit der Kontrolle des Zugangs und der Setzung von Regeln und Standards entfalten Plattformen eine Macht, die politisch und gesellschaftlich immer noch kaum reguliert ist. Es lohnt sich also, die „Macht der Plattformen“ (Seemann 2020) genauer zu untersuchen, um besser verstehen zu können, auf welches Resonanzfeld Hassakteure und extremistische Bewegungen bauen können und inwieweit sie von medialen Gatekeepern wahrgenommen werden.

Mit der Kontrolle des Zugangs und der Setzung von Regeln und Standards entfalten Plattformen eine Macht, die politisch und gesellschaftlich immer noch kaum reguliert ist.

Dieser Beitrag möchte einen Überblick darüber schaffen, welche Rolle die Gestaltung von Plattformen für die Verbreitung von rechtsextremen und verschwörungsideologischen Inhalten spielt. Diese Frage speist sich aus dem Interesse zu zeigen, wie ein zunehmend privatwirtschaftlich organisierter

Umgang mit digitalem Hass und Extremismus die Einwirkungsmöglichkeiten der Zivilgesellschaft und Politik beeinträchtigt bzw. neu ausrichtet. Aber auch die Plattformen ändern sich: Denn – so argumentieren wir – am Umgang mit Hass und Extremismus lassen sich größere Muster der Moderation und Verwaltung von digitalen Plattformen ablesen. Um dieses Argument zu systematisieren, gehen wir in vier Schritten vor: Zunächst geben wir eine kurze Einführung in die Politik digitaler Plattformen und nähern uns daraufhin deren Bedeutung für digitalen Hass an. Auf Grundlage eines von uns gesammelten Datensatzes von über 400 Dokumenten identifizieren wir Determinanten der Plattformpolitik und schlagen eine Periodisierung der Maßnahmen vor, die ein besseres Verständnis zum Umgang von Plattformen mit Hass und Extremismus geben.

Plattformpolitik: a primer

Digitale Plattformen sind politische Akteure, die sich an der Aushandlung von öffentlichen und privaten Werten und Interessen beteiligen. Sie tun dies u. a. durch diverse Designelemente, die ein bestimmtes Verhalten anhand von Metriken belohnen oder sanktionieren: Hierzu zählen die Anzahl von Likes, Viewzahlen und Reposts, die sich als Seismografen für die Wertigkeit oder Relevanz von Aussagen in der Wahrnehmung der

Digitale Plattformen sind politische Akteure, die sich an der Aushandlung von öffentlichen und privaten Werten und Interessen beteiligen.

Öffentlichkeit etabliert haben (Mau 2017). Zudem bieten digitale Plattformen die Möglichkeit der Monetarisierung von Inhalten, beispielsweise durch Gewinnbeteiligung oder indirekte Werbung. Dies hat einen unüberschaubaren Markt von digitalen Angeboten geschaffen, die personalisiert aufbereitet werden und zum Verweilen auf der Plattform bzw. zur Interaktion mit anderen Nutzer*innen animieren.

Sichtbarkeit und Reichweite sind die zentralen Ressourcen im Ringen um digitale Aufmerksamkeit. Doch sind die Voraussetzungen zur Erlangung dieser durchaus unterschiedlich und hängen von verschiedensten Regelungen und deren (selektiver) Durchsetzung ab. Schließlich behalten sich Plattformen vor, gegen Nutzer*innen vorzugehen, die gegen ihre Nutzungsbedingungen verstoßen. Hierbei werden allerdings Unterschiede deutlich: Wird man von der Plattform als eine Person mit Nachrichtenwert (Newsworthiness) identifiziert, erhöht sich die Sichtbarkeit, während gleichzeitig die moderativen Eingriffe verringert werden (DiResta und Debutts 2021). So konnte Trump über Jahre hinweg rassistische und muslimfeindliche Beiträge absetzen. Andere Nutzer*innen der Plattform wären für solche Beiträge längst von der Plattform verwiesen worden. Beispielsweise forderte Trump 2015 ein prinzipielles Einreiseverbot für Muslim*innen in die USA (Frenkel und Kang 2021, 23f.).

Es ist eine ungemein politische Frage, wie Plattformen ihre Regeln gestalten und anwenden. Welche gesellschaftlichen und kommerziellen Kräfte sich an der Aushandlung der Regeln beteiligen, wie Plattformen sich selbst reformieren oder von Staaten reguliert und zu Veränderungen bewegt werden, untersucht die Plattform Governance Forschung (Flew und Martin 2022). Der Fokus liegt auf den Regeln und der Praxis der Moderation von Inhalten, welche den Charakter einer Plattform – und damit auch die Art der Interaktion ihrer Nutzer*innen – maßgeblich bestimmen (Gillespie 2018, 5). Über die Definition dessen, was legitim auf einer Plattform geäußert werden kann und die Möglichkeit, Akteure vom Diskurs auszuschließen, üben Plattformbetreiber politische Macht aus. Als gewinnorientierte Unternehmen stellen sie global geltende Regeln auf, die nicht selten durch Nutzer*innen, Regierungen und Zivilgesellschaften herausgefordert werden (Gorwa 2019a).

Inhaltsmoderation ist in Hinblick auf *Plattformpolitik* wohl am wahrnehmbarsten, insofern die Maßnahmen – bspw. die Löschung, Sperrung, Reichweitenreduzierung, Warnhinweise und Demonetarisierung – einen Ausdruck auf der Plattform finden. Zugleich gilt die konkrete Klick-Arbeit der tausenden Inhaltsmoderator*innen als psychisch belastende Tätigkeit, weshalb sie oft in Niedriglohnländer ausgelagert wird (Gray und Suri 2019).

Inhaltsmoderation ist in Hinblick auf Plattformpolitik wohl am wahrnehmbarsten, insofern die Maßnahmen – bspw. die Löschung, Sperrung, Reichweitenreduzierung, Warnhinweise und Demonetarisierung – einen Ausdruck auf der Plattform finden.

Die (exekutiven) Moderationsteams setzen die Richtlinien um, die durch die (legislativen) Content-Policy-Abteilungen der jeweiligen Plattformen verfasst werden. Dieses *Policy-Making* umfasst Jurist*innen, Politolog*innen und praxisnahe Expert*innen der Unternehmen, die zur Festsetzung von Richtlinien zusammenkommen. Unter Hinzuziehung externer Expertisen (*Stakeholder Engagement*) wird auf neue gesellschaftliche Entwicklungen reagiert und bestehende Richtlinien werden evaluiert.

Eine weitere Säule der Plattformpolitik umfasst die Aufsicht (*Oversight*) über die Regel(um)setzung. Hierzu zählen staatliche Regulierungsbehörden, hybride Kontrollgremien und eigens eingerichtete Räte, die eine Form der Kontrolle über die Unternehmen ausüben sollen. Am bekanntesten für letztere ist das Oversight Board von Meta. Die 24 Expert*innen aus den Bereichen Jura, Journalismus und Menschenrechte, die vom Konzern selbst nominiert wurden, haben die Aufgabe, bei kritischen Entscheidungen ein bindendes Urteil zu fällen. Staatliche Behörden nehmen über Gesetze wie das Netzwerkdurchsetzungsgesetz (NetzDG) Einfluss auf die Plattformpolitik. Dagegen ist eine demokratische Aufsicht in Form von Plattformräten – umfassend aus verschiedenen Sektoren der Gesellschaft – noch kaum abzusehen (Kettmann und Fertmann 2022).

Schließlich gibt es in jedem Konzern *Public Policy*-Abteilungen, die für Außendarstellungen und Stellungnahmen gegenüber der Öffentlichkeit zuständig sind. Sie sind verantwortlich für Lobby-Arbeit und die Beziehungspflege zu *trusted partners*. Sie treten auf Veranstaltungen und in Parlamenten auf, schreiben in Zeitschriften und auf Blogs und stehen der Presse für Interviews zur Verfügung.

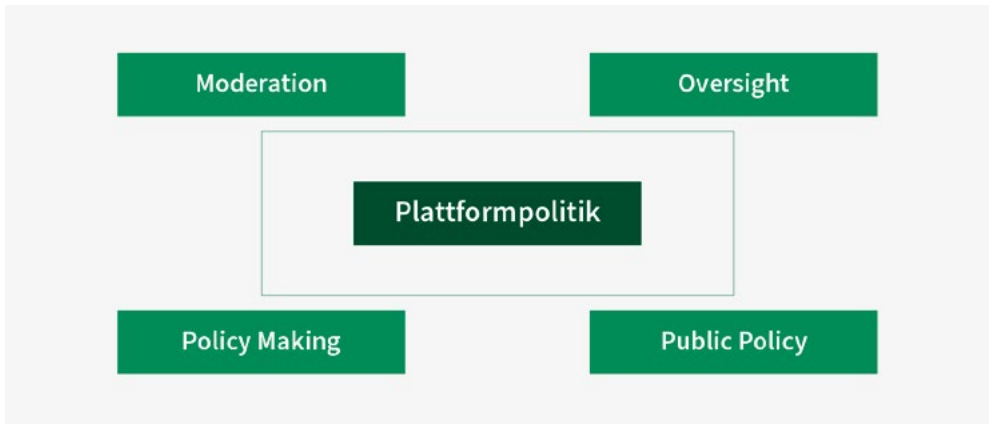


Abbildung 1: Die vier Säulen der Plattformpolitik

Plattformpolitik als ein Segment von Plattform-Governance verstehen wir als den Prozess der Aushandlung von Positionen und Werten sowie die Steuerung der Sichtbarkeit und Aufmerksamkeit von Nutzer*innen auf einer – oder mehreren – Plattform(en). Plattformpolitik vollzieht sich auf verschiedenen Ebenen: in geschlossenen Expert*innengruppen, in den Aufsichtsgremien, in zivilgesellschaftlichen Stakeholder-Audits, aber eben auch in öffentlichen Diskursen und der *public relations* eines jeden Plattformunternehmens.

Plattformpolitik als ein Segment von Plattform-Governance verstehen wir als den Prozess der Aushandlung von Positionen und Werten sowie die Steuerung der Sichtbarkeit und Aufmerksamkeit von Nutzer*innen auf einer – oder mehreren – Plattform(en).

Während viele Entscheidungsfindungsprozesse intransparent ablaufen (Pasquale 2015), werden Unternehmen heute verstärkt zur Positionierung und Rechtfertigung von Maßnahmen gebracht. Dies trifft insbesondere auf gesellschaftlich relevante Themen wie den Umgang mit digitalem Hass und Extremismus zu, bei denen digitalen Plattformen eine zentrale Rolle zugeschrieben wird.

Hass, (Rechts-)Extremismus und Plattformpolitik

Noch vor wenigen Jahren konnten Hakenkreuze und Holocaustleugnung unter dem Deckmantel der Meinungsfreiheit auf Plattformen Verbreitung finden und verschwörungsideologische Gruppen konnten ihren digitalen Aktivismus problemlos monetarisieren (Marantz 2020). Hass und Hetze gegen Minderheiten haben über digitale Plattformen eine solche Reichweite erlangt, dass ein Großteil der Bevölkerung schon einmal damit konfrontiert wurde (Geschke et al. 2019). Dies führt dazu, dass insbesondere weniger geschützte Gruppen sich aus öffentlichen Diskursen zurückziehen und somit (noch) weniger sichtbar werden. Die dominanten Stimmen bilden somit bei Weitem nicht moderne Gesellschaften ab – und hier sind noch nicht einmal die Verstärkereffekte algorithmischer Diskriminierung angesprochen sind (Noble 2022).

Mit der Verschiebung politischer Diskurse in den digitalen Raum haben sich die Regeln und Bedingungen des Austausches zwischen Gruppen und Individuen geändert. Plattformen wurden Austragungsort gesellschaftlicher Konflikte, die – angetrieben durch auf Interaktion getrimmte Algorithmen – nicht selten in Anfeindungen enden, Antagonismen verfestigen und ein

Mit der Verschiebung politischer Diskurse in den digitalen Raum haben sich die Regeln und Bedingungen des Austausches zwischen Gruppen und Individuen geändert. Plattformen wurden Austragungsort gesellschaftlicher Konflikte.

Verständnis für andere Positionen unterminieren. Wir finden eine Form der affektiven Polarisierung vor, die gerade durch die emotional geführte Interaktion mit gegenteiligen Meinungen die eigene Weltsicht nur weiter bestärkt (Törnberg 2022). Radikale und extremistische Akteure, die an der Auflösung von verständnisorientierten, offenen Diskursen interessiert sind, nutzen diese strukturellen Vorteile und genießen weitgehende Freiheiten, ihre Ideen und Vorurteile zu verbreiten.

Auch entwickelten sich über digitale Plattformen „neue“ Online-Subkulturen, die offen zu Hass und zu Gewalt gegen Minderheiten oder politische Gegner*innen aufrufen. Aus ihnen heraus rekrutierten sich Islamisten und rechtsextreme Gewalttäter, die pluralistische Gesellschaften mit terroristischen Mitteln erschüttern wollten. Zugleich ließ sich eine Tendenz der Selbst-Aktivierung von digital affinen Menschen beobachten, die losgelöst von Aufrufen und Organisationen schwer vorhersagbare Gewaltakte durchführen, die wiederum zum Nachmachen animieren sollen (Fielitz 2022). Der Anschlag in Christchurch im März 2019 setzte neue Maßstäbe, weil er der erste war, der ein digitales Publikum via Livestream an dem Mord an Dutzenden von Menschen teilhaben ließ. Auch weil Teile des Livestreams allein auf Facebook mehr als 1,5 Millionen mal von Nutzer*innen hochgeladen wurden, mussten Plattformen und Politik ein Zeichen setzen (Macklin 2019). Mit dem

Christchurch-Call verpflichteten sich sodann Staat und Plattformunternehmen dazu, die Möglichkeit zu unterbinden, soziale Medien für die Organisation und Förderung von Terrorismus und gewalttätigem Extremismus zu nutzen (Christchurch Call 2019).

Mit der Zeit entwickelte sich der Umgang mit digitalem Hass und Extremismus von einem Randphänomen zu einer wesentlichen Herausforderung der Plattformpolitik – insbesondere seitdem sich soziale Medien zu einem zentralen Mobilisierungsraum entwickelt haben und Personen ihre Reichweite nutzen, um direkt oder indirekt über die Plattformen Menschengruppen abzuwerten oder zu Gewalt aufzurufen. Im deutschen Kontext wurde mit dem NetzDG 2017 ein weltweit erster Rechtsrahmen geschaffen, der Plattformen in die juristische Verantwortung nimmt, wenn sie nicht schnell genug gegen strafrechtlich relevante Hass-Postings vorgehen (Gorwa 2021). Es folgte eine restriktive Phase von Account-Sperrungen, das sogenannte Deplatforming (Fielitz und Schwarz 2020), in der Accounts, die gegen die Nutzungsregeln verstießen, der Zugang verwehrt wurde. Diese Deplatforming-Welle hat in den Jahren der Covid19-Pandemie neuen Schwung erhalten und bedarf infolgedessen weniger Rechtfertigung der Plattformunternehmen (Magalhães und Katzenbach 2020).

Determinanten der Plattformpolitik

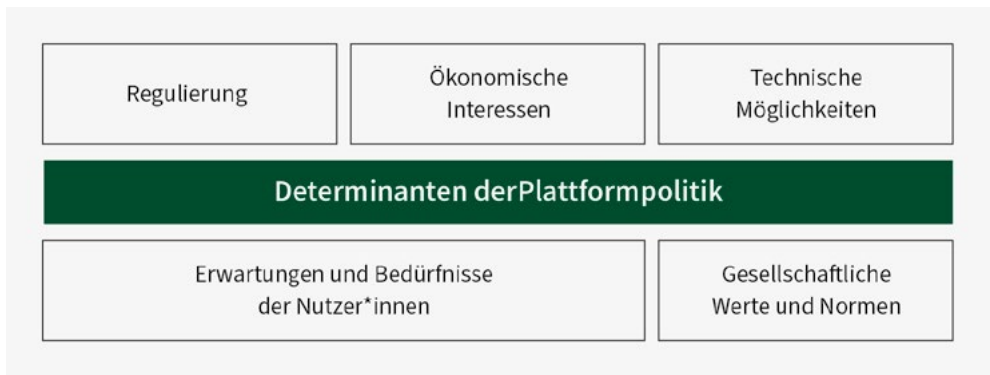


Abbildung 2: Determinanten der Plattformpolitik

Wie bereits angedeutet, ergeben sich Maßnahmen im Umgang mit Hass und Extremismus aus dem Zusammenspiel verschiedener Faktoren, die die wechselseitige Bedingtheit von Kontext, Ereignis und Intervention bestimmen. Wir beleuchten dabei fünf Determinanten: 1. Regulierung, 2. Ökonomische Interessen, 3. Erwartungen und Bedürfnisse der Nutzer*innen, 4. Technische Möglichkeiten, 5. Gesellschaftliche Werte und Normen.

Eine zentrale Determinante für Plattformpolitik ist die *Regulierung*, also: „rechtlich-regulative Eingriffe in die soziale Regelungshoheit der Plattformen“ (Dolata 2020, 26). Plattformen sind an externe

rechtliche Rahmenbedingungen wie Datenschutzgesetze, Urheberrechtsgesetze und nationale Gesetze gebunden. Am relevantesten ist das erwähnte NetzDG oder der ab 2024 europaweit in Kraft tretende Digital Service Act. Diese Gesetze sollen Plattformen zur Umsetzung von Standards bewegen und können bei Zuwiderhandlung zu Geldstrafen führen (Gorwa 2019a). Weder Regierungen noch supranationale Zusammenschlüsse von Regierungen wie die Europäische Union oder die Vereinten Nationen haben es bislang geschafft, einen bindenden Rechtsrahmen zu verabschieden, der die Macht der transatlantisch operierenden Plattformen umfassend reguliert. Darüber hinaus ist eine Co-Governance auszumachen, indem die rechtlichen Rahmenbedingungen durch freiwillige Standards, transnationale Regeln und Aufsichtsgremien ergänzt werden. Hierbei ist die aktive Rolle privater Akteure herauszustellen, die ihren eigenen Handlungsrahmen mitregulieren (Gorwa 2019b).

Als Beispiele für Institutionen, in denen freiwillige Standards und transnationale Regeln ausgearbeitet werden, können das 2017 gegründete *Global Internet Forum to Counter Terrorism (GIFCT)* oder der *Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online* herangezogen werden. Ersteres ist eine Initiative der Plattformindustrie, um Terrorismus über den Austausch an Informationen und Technologien zu verhindern. Zweiteres wurde nach den rechtsterroristischen Attentaten auf zwei Moscheen im März 2019 vom neuseeländischen Außenministerium initiiert. Dabei handelt es sich um eine allgemeine und unverbindliche Verpflichtung zur Verhinderung der Verbreitung terroristischer Inhalte im Internet. Der Christchurch Call wurde damals von 17 Ländern und 8 Unternehmen (darunter Facebook, Google, Twitter) erstunterzeichnet. Die Unterzeichnung war auch ein Versprechen für weitere informelle Zusammenarbeiten zwischen Regierungen und Unternehmen (Douek 2019). Das bekannteste Beispiel für ein *Aufsichtsgremium* ist das bereits erwähnte Oversight Board von Meta, das sich mit der Überwachung, Moderation und Bewertung von Inhalten auf den Meta-Plattformen befasst. Das mit externen Expert*innen besetzte Gremium soll unabhängige Einschätzung zu strittigen Inhalten sowie Vorstellungen für Gegenmaßnahmen einleiten und Standards manifestieren (Harris 2020). Mit Blick auf derartige Aufsichtsgremien kann von quasi-hoheitlichen Parallelstrukturen zu den demokratisch legitimierten gesellschaftlichen Institutionen gesprochen werden (Dolata 2020).

Eine zweite Determinante bei der Ausrichtung von Plattformpolitik ist das *ökonomische Interesse* der Plattformunternehmen. Konjunkturzyklen, Inflation oder die Arbeitsmarktsituation können politische Entscheidungen beeinflussen, denn die Monetarisierung von Inhalten, Abonnements, Werbung oder Kooperationen sind zentrale Elemente im Geschäftsmodell von Plattformunternehmen und wichtig, um wettbewerbsfähig zu bleiben oder die eigene Marktposition zu stärken (Barwise und Watkins

Eine zweite Determinante bei der Ausrichtung von Plattformpolitik ist das ökonomische Interesse der Plattformunternehmen.

2018). Bspw. hat der Gewinn digitaler Anzeigen bei Meta 2022 einen erheblichen Einbruch erlitten, sodass das Unternehmen einen Wertverlust von 70 % zu verbuchen hatte und im November desselben Jahres einen Stellenabbau von 13 % verkündete, der auch Policy-Strukturen im Unternehmen betrifft (Jaspert 2023). Ein erneuter Wahlkampf von Trump könnte dem Einbruch nämlich entgegenwirken: Die Vergangenheit hat gezeigt, dass das Trump-Team allein zwischen April und Oktober 2020 rund 145,1 Mio. US-Dollar für Werbung bei Meta ausgab. Somit ist die erneute Freischaltung der Accounts von Donald J. Trump nicht zuletzt im Lichte ökonomischer Interessen zu betrachten.

Neben der Regulierung und den ökonomischen Interessen sind auch die *Erwartungen und Bedürfnisse* der Nutzer*innen eine wichtige Determinante für das politische Handeln von Plattformen. Aus diesem Grunde haben so gut wie alle digitalen Plattformen ein Flagging bzw. Reporting System als Kennzeichnungsmechanismus für schädliche oder unliebsame Inhalte integriert. Dieser Governance-Mechanismus dient

Neben der Regulierung und den ökonomischen Interessen sind auch die Erwartungen und Bedürfnisse der Nutzer*innen eine wichtige Determinante für das politische Handeln von Plattformen.

der Bewältigung und Kuratierung großer Mengen nutzergenerierter Inhalte. Zudem fungiert er als Rechtfertigung für die Plattformunternehmen, wenn diese beschließen, Inhalte oder Accounts zu löschen oder die Sichtbarkeit zu reduzieren (Crawford und Gillespie 2016). YouTube schrieb 2020 auf seinem Unternehmensblog über die Bedeutung des Flagging Systems: „But no system is perfect, so we make sure if you see something that doesn't belong on YouTube, you can flag it for us and we'll quickly review it“ (The YouTube Team 2020). Erwartungen und Bedürfnisse der Nutzer*innen werden teilweise auch über Umfragen auf den Plattformen abgefragt. Facebook hat zwar Abstimmungen zu politischen Änderungen als mögliches Modell schnell verworfen (Gillespie 2018), dagegen nutzt Twitter regelmäßiger Umfragen. Es wurde beispielsweise unter dem Hashtag #TwitterPolicyFeedback eine öffentliche Resonanz zu Policy-Änderungen eingeholt. Öffentliche Umfragen gibt es auch nach der Übernahme von Elon Musk weiterhin. Diese sind inzwischen jedoch meistens PR-Aktionen und das Ergebnis wird nicht mehr zwingend umgesetzt. Fun-Fact: Ende 2022 ließ Musk in einer Abstimmung sogar über die eigene Chefposition abstimmen, bei der er verlor. Der Großteil der Nutzer*innen stimmte für den Rücktritt des Milliardärs. Nicht aufgrund der Abstimmung, sondern vielmehr aus Kapazitätsgründen trat Musk dann im Mai 2023 tatsächlich als CEO zurück und setzte die Werbemanagerin Linda Yaccarino als neue CEO der X Corp. (zuvor Twitter) ein.

Technologische Beschränkungen und *neue bzw. weiterentwickelte technische Möglichkeiten* beeinflussen ebenfalls die Politik der Tech-Konzerne. Allen voran sind Fortschritte im Bereich der Künstlichen Intelligenz zu nennen, auf die sich Plattformen seit 2017 immer öfter beziehen, wenn sie politische

Entscheidungen treffen und umsetzen. YouTube zum Beispiel verkündete: „[w]e will use our cutting-edge machine learning more widely [...] to flag violent extremist content for human review and we've seen tremendous progress“ (Wojcicki 2017). Seit der Corona-Pandemie hat sich der Automatisierungsgrad ein weiteres Mal erhöht (Magalhães und Katzenbach 2020). Facebook resümierte 2021, dass es knapp 98 Prozent der Hass-Inhalte entfernt habe, ohne dass es darauf hingewiesen worden sei, während die Quote 2017 noch bei 24 Prozent lag (Facebook 2021). Mehr noch als das Entfernen von Inhalten ist eine Tendenz zur Einschränkung festzustellen. Hier werden Inhalte anhand von bestimmten Wortkombinationen und Sentimentanalysen als wichtig oder weniger wichtig bewertet und entsprechend sichtbarer oder unsichtbarer gemacht. Einschränkungen haben den Vorteil für Plattformen, das der sogenannte „borderline content“ auf den Plattformen verbleiben kann, ohne große Reichweite zu erzielen (Gillespie 2022). Das hat aber auch den Nachteil für Nutzer*innen, dass sie über die Einschränkungen (bislang) nicht informiert werden (Savolainen 2022).

Schließlich sind *gesellschaftliche Werte und Normen* eine relevante Determinante, die auch auf Plattformpolitik einwirkt. Diese werden vor allem von zivilgesellschaftlichen Gruppen, Nichtregierungsorganisationen, Wissenschaftler*innen, Investor*innen oder

Gesellschaftliche Werte und Normen sind eine relevante Determinante, die auch auf Plattformpolitik einwirkt.

Einzelpersonen über Lobbyarbeit, Kampagnen oder Debatten an die Plattformen herangetragen (Gorwa 2019b). Als Beispiele können #MeToo oder #Black Lives Matter genannt werden, im Zuge derer Sexismus bzw. Rassismus in der Plattformökonomie stärker in das Blickfeld rückten. YouTube schrieb nach dem Tod von George Floyd: „wir setzen alles daran, systematischen Rassismus aufzudecken und endgültig zu besiegen“ (Wojcicki 2020). Bei der Betrachtung von gesellschaftlichem Einfluss auf die Plattformpolitik sind auch Ambivalenzen festzustellen. Eine entscheidende Ambivalenz zeigt sich in den Positionen zum Umgang mit Hate Speech, kultureller Sensibilität und politischen Desinformationen auf der einen Seite und dem Zugang zu Informationen und der Meinungsfreiheit auf der anderen Seite (Maddox und Malson 2020).

Paradigmen der Plattformpolitik im Wandel

Im zeitlichen Verlauf sind verschiedene Paradigmen der Plattformpolitik auszumachen, die entlang einer Periodisierung einzelner Phasen idealtypisch unterschieden werden können. Hierzu schlagen wir eine Periodisierung vor, die sich aus einer triangulierten Analyse verschiedener Datenquellen ergibt. Als Datenquellen wurden die Terms of Services, Community Guidelines und Blogbeiträge von Twitter, Meta (früher Facebook) und YouTube herangezogen. Periodisierungen als Heuristiken können temporäre Entwicklungen nachzeichnen und verschiedene Zeitabschnitte kontrastieren.

Die Grenzen der Zeitabschnitte sind dabei nicht scharf zu ziehen oder gänzlich abgeschlossen, allerdings lassen sich charakterisierende Merkmale ableiten, die die Handlungslogiken von Plattformen besser verstehen lassen.



Abbildung 3: Periodisierung von Paradigmen der Plattformpolitik

In der frühen Phase, von der Liveschaltung der Plattformen bis etwa 2010, fanden die ersten Schritte der Plattformmoderation statt, aus denen sich Policies entwickelten, die mit der Zeit mehr und mehr institutionalisiert wurden. Anfangs waren die Terms of Services und Community Guidelines recht einfach strukturiert. Alles war erlaubt, solange kein Copyright verletzt oder Gewalt, Belästigungen sowie Pornografie geteilt wurde. Wurde gegen diese Regeln verstoßen, löschten die Plattformen die Inhalte, wiederholte sich der Verstoß, wurde das Profil gelöscht. Die Moderationskapazitäten waren im Vergleich zu heute sehr gering und durch die wenig standardisierten Abläufe wurde oft erst auf offensichtliche Verstöße reagiert. Der Fokus in dieser Phase lag vor allem auf dem Verhalten einzelner Individuen und weniger auf gruppenbezogenen Dimensionen. Dieser Fokus prägte ebenso das fallspezifische Handeln der Plattformen, indem Erfahrungen zu Richtlinien wurden. Das Policy Making wurde 2008 von Youtube wie folgt beschrieben: „As the YouTube Community expands and evolves, we’re always trying to keep pace by creating policies

that reflect innovative new uses of YouTube and the diverse content posted by users every day” (The Youtube Team 2008). In dieser Phase war das Selbstregulierungsparadigma stark, wonach die Konzerne vorgaben, Probleme auf ihren Plattformen selbstständig zu lösen.

Die zweite Phase beginnt etwa 2010 und reicht bis ins Jahr 2015. Diese Phase ist sowohl durch eine Zunahme politischer Inhalte auf den Plattformen als auch durch mehr Transparenz in Entscheidungsfindungen gekennzeichnet. Die Politisierung lässt sich am Arabischen Frühling zeigen. Hier nahmen digitale Plattformen erstmals eine zentrale Rolle in der Mobilisierung politi-

Die zweite Phase beginnt etwa 2010 und reicht bis ins Jahr 2015. Diese ist sowohl durch eine Zunahme politischer Inhalte auf den Plattformen als auch durch mehr Transparenz in Entscheidungsfindungen gekennzeichnet.

scher Kräfte ein. Mit der Politisierung einher ging außerdem ein Anstieg des Online-Extremismus. Radikale politische Akteure nutzten verstärkt das Potenzial, sich zu vernetzen, Informationen auszutauschen und sich und andere zu radikalisieren. Darauf reagierten die Plattformunternehmen, indem sie die Nutzer*innen mit in die Verantwortung nahmen. Twitter schrieb beispielsweise dazu, „[y]ou can help out, too, by reporting and blocking“ (Twitter 2012). Die Nutzer*innen sollten also fortan helfen, Inhalte und Accounts zu melden, damit die Plattformen schädliche Inhalte schneller löschen können. Zugleich wurde das Löschen von Inhalten erstmals zum Politikum, nachdem YouTube 2012 ein anti-muslimisches Video löschte, das nicht gegen die Richtlinien verstieß. Ausschlaggebend waren mehrere arabische Länder, die den Zugang zu YouTube blockierten, damit das Video keine Verbreitung fand. Diese Maßnahme wurde stellenweise jedoch als Zensur der Meinungsfreiheit aufgefasst (Yorck 2012). Die Rufe nach mehr Transparenz wurden lauter und so kam es in dieser Phase ebenfalls zu den ersten Veröffentlichungen von Transparenzberichten (Kessel 2012). Dementsprechend ist für diese Phase einerseits ein Nutzerverantwortungsparadigma festzustellen, da Nutzer*innen mehr und mehr in die Inhaltsmoderation einbezogen wurden, und andererseits ist ein Transparenz- und Rechenschaftsparadigma festzustellen, da Plattformen Moderationspraktiken und Entscheidungsprozesse nun offensiver kommunizierten.

Schon vor der Gründung des Islamischen Staats im Juni 2014 ging die Entwicklung dahin, dass Plattformen von terroristischen Gruppen zur Rekrutierung neuer Mitglieder genutzt wurden. Auch Gewaltdarstellungen waren in dieser Zeit omnipräsent. Als es dann im November 2015 zu den Terroranschlägen in Paris kam, gerieten die Plattformunternehmen unter starken Druck und verstärkten als Reaktion die Zusammenarbeit untereinander sowie mit Regierungen und zivilgesellschaftlichen Institutionen. Das gemeinsame Ziel: Radikalisierung und Hate Speech bekämpfen. Dafür wurde zunächst das EU-Internet-Forum zum Austausch ins Leben gerufen und im Jahre 2017 mit dem schon

erwähnten GIFCT auf eine globale Ebene gehoben (Facebook 2017). Im Zuge dieser Kooperationen begannen Plattformen sogenannte Hashes auszutauschen, das sind digitale Fingerabdrücke extremistischer Inhalte, damit andere Plattformen diese Inhalte ebenfalls entfernen konnten. Außerdem wurden nach der US-Wahl 2016 und den Vorwürfen der ausländischen Einflussnahme durch Desinformationen die Investitionen in die Technologie erhöht, um mit Künstlicher Intelligenz und maschinellem Lernen automatisch nach Inhalten zu suchen und diese schneller zu identifizieren. Der rechtsterroristische Anschlag in Christchurch 2019 heizte die Debatte hinsichtlich des Umgangs mit extremistischen Inhalten ein weiteres Mal an. Auch hier wurde mit einer neuen Kooperation, dem Christchurch Call, reagiert und Künstliche Intelligenz als die zeitgemäße Maßnahme zur Bekämpfung schädlicher Bilder und Videos stark gemacht. Die Paradigmen der Kooperation und Technikzentrierung sind in dieser Phase allgegenwärtig.

Die Paradigmen zeigen sich in der vierten Phase ebenso dynamisch. Die COVID-19-Pandemie führte 2020 weltweit zu einer erneuten Zunahme an Online-Aktivitäten, da sich das Soziale stark in den digitalen Raum verlagerte. Im Zuge dessen kam es zu einer Hochkonjunktur an Desinformationen und neuen Ausprägungen von Online-Extremismus. Nach der Erstürmung des US-Kapitols

Aktuell ist das Paradigma der Meinungsfreiheit stärker und es wird postuliert, dass die Grenzen der Meinungsfreiheit nicht durch Plattformen festgelegt werden sollten. Außerdem ist ein Regulierungsparadigma festzustellen.

im Januar 2021 wurde der abgewählte US-Präsident Donald J. Trump von Plattformen wie Facebook, Twitter und YouTube gesperrt. Zuvor hatte er auf diesen Plattformen zum „Save America March“ aufgerufen, von dessen Teilnehmenden die Gewalt ausging. Auch im Nachgang distanzierte sich Trump nicht von den Geschehnissen. Knapp zwei Jahre nach der Sperrung wurden Trumps Accounts bei YouTube, Meta und X wieder freigeschaltet. Dem vorgelagert war eine Debatte über Meinungsfreiheit, schädliche Inhalte, die entfernt werden sollten, geschmacklose Inhalte, die eine liberale Demokratie aushalten müsse, und die Macht privater Unternehmen, Strafen auszusprechen, während demokratische legitimierte Institutionen keine Strafen verhängen würden (Clegg 2023). Elon Musk propagierte nach seiner Twitter-Übernahme, dass Trump nicht gegen die Twitter-Richtlinien verstoßen habe und ließ den Ex-Präsidenten freischalten. Meta zog in diesem Belang die Expertise des neu gegründeten Oversight Boards heran und kam letztlich ebenfalls zu dem Ergebnis, Trump freizuschalten. Darüber hinaus wurde das Höchstmaß einer zweijährigen Sperrung festgeschrieben. Facebook gab in diesem Prozess zu Protokoll: „[t]he Oversight Board is not a replacement for regulation, and we continue to call for thoughtful regulation in this space“ (Clegg 2021) und fordert damit mehr Regulierung der eigenen Handlungsmacht. Aktuell ist das Paradigma der Meinungsfreiheit stärker und es wird postuliert, dass die Grenzen der Meinungsfreiheit nicht durch Plattformen festgelegt werden sollten. Außerdem

ist ein Regulierungsparadigma festzustellen: Es wird ein klarerer Rechtsrahmen gefordert, der den Handlungsraum der Plattformen einschränkt. Ein spätes Eingeständnis, dass die Plattformindustrie Hass und Online-Extremismus nicht federführend lösen kann.

Ausblick: Schnellebigkeit der Plattformpolitik

Der Beitrag gibt einen Überblick über Plattformpolitik mit einem Fokus auf den Umgang mit Hass und Extremismus. Darüber hinaus wurden die Einflussfaktoren und Paradigmen von Plattformpolitik im zeitlichen Verlauf von mehr als 15 Jahren abgezeichnet. Es wurde aufgezeigt, dass sich sowohl die Plattformpolitik als auch die gesellschaftliche Rolle digitaler Plattformen in einem permanenten Wandel befinden. Fragen der Inhaltsmoderation als Kern von Plattformpolitik haben sich von einem eher sekundären Service von Plattformen hin zu einem zentralen Kuratierungsmechanismus entwickelt, der den Nutzer*innen Transparenz und Sicherheit suggerieren soll. Zugleich zeigt sich, wie andere Schutzmechanismen von Plattformen verschwinden, z. B. das Blocken von Accounts, der Zugang zu Schnittstellen für die Forschung eingeschränkt wird und Narrative und Akteure (wieder) Zulauf gewinnen, die einst von den Plattformen verwiesen wurden.

Die Schnellebigkeit der Plattformpolitik macht es daher schwer, Prognosen über zukünftige Entwicklungen zu geben. Prinzipiell treffen zwei widersprüchliche Entwicklungen aufeinander: auf der einen Seite globale Gesellschaften, die sensibler geworden sind für digitale Diskriminierungen und Gewalt, und auf der anderen Seite Konzerne, die kalkulieren, wie viel Einschränkungen die Gewinnmarge beeinträchtigen könnten. Zwar gibt es in der Europäischen Union stärkere Bestrebungen, Kontrollmechanismen für Plattformen zu installieren. Für die Zivilgesellschaft ist die Macht der Plattformen bisher aber nur indirekt beeinflussbar. Es gibt mehr Einbindungen durch Feedbackforen oder die Einholung von Expertisen. Dennoch entstehen Richtlinien und Maßnahmen in einer kaum nachvollziehbaren Art und Weise und werden oft durch Sachzwänge gerechtfertigt.

Was in den vergangenen Jahren allerdings gelungen ist: den Mythos einer politischen Neutralität von Plattformen zu brechen. Es gibt heute ein viel größeres Bewusstsein für die politischen Auswirkungen des (Nicht-)Handelns von Plattformen und eine kritische Begleitung der Gestaltungsmacht der Plattformunternehmen. Denn: Ob gewollt oder nicht, betreiben die Konzerne Plattformpolitik – in Form der Organisierung ihrer Machtbasis und in Form der Einflussnahme auf politische Systeme und die transnationale Nutzer*innengemeinschaft. Dies hat Auswirkungen darauf, wie sich Akteure und Ideologien verbreiten und eine breite Masse erreichen. Für

Was in den vergangenen Jahren gelungen ist: den Mythos einer politischen Neutralität von Plattformen zu brechen.

Vertreter*innen des Rechtsextremismus und von Verschwörungsideologien bedeuten die jüngsten Volten der Plattformpolitik, dass sie unverhofft von ihren Alternativplattformen in den digitalen Mainstream zurückkehren können und somit Ressourcen zur Reichweitengenerierung nutzen können, die ihnen temporär verwehrt blieben.

Angesichts der Rolle, die insbesondere soziale Medien beim Aufstieg rechtsextremer Parteien und beim Umsichgreifen rassistischer Gewalt gespielt haben, sollten die Entwicklungen der Plattformpolitik aufmerksam von der Zivilgesellschaft verfolgt werden.

Maik Fielitz verantwortet den Bereich Rechtsextremismus- und Demokratieforschung am IDZ. Er ist des Weiteren wissenschaftlicher Referent am IDZ und Ko-Leiter der Forschungsstelle der Bundesarbeitsgemeinschaft „Gegen Hass im Netz“. Er leitet das Forschungsprojekt „Die Politik digitaler Plattformen im Kontext rechtsextremer und verschwörungsideologischer Mobilisierung“ im Rahmen des BMBF-Forschungsverbunds NEOVEX.

Marcel Jaspert ist wissenschaftlicher Mitarbeiter am IDZ und forscht im Zuge des NEOVEX Verbundprojekts zur Rolle und den Rechtfertigungsmustern von digitalen Plattformen bei der Verbreitung und Einhegung rechtsextremer und verschwörungsideologischer Dynamiken.

Literatur

- Barwise, Patrick/Watkins, Leo (2018). The Evolution of Digital Dominance: How and Why We Got to GAFA. In: Moore, Martin / Tambini, Damian (Hg.). Digital Dominance. The Power of Google, Amazon, Facebook, and Apple. Oxford, Oxford University Press, 21–49.
- Bond, Shannon (2023). YouTube will no longer take down false claims about U.S. elections. Hg. v. NPR. Online verfügbar unter <https://www.npr.org/2023/06/02/1179864026/youtube-will-no-longer-take-down-false-claims-about-u-s-elections> (abgerufen am 15.07.2023).
- Christchurch Call (2019). The Christchurch Call to Action To Eliminate Terrorist and Violent Extremist Content Online. Online verfügbar unter <https://www.christchurchcall.com/assets/Documents/Christchurch-Call-full-text-English.pdf>.
- Clegg, Nick (2021). In Response to Oversight Board, Trump Suspended for Two Years; Will Only Be Reinstated if Conditions Permit. Facebook Blog. Online verfügbar unter <https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/> (abgerufen am 14.07.2023).
- Clegg, Nick (2023). Ending Suspension of Trump's Accounts With New Guardrails to Deter Repeat Offenses. Facebook Blog. Online verfügbar unter <https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/>, (abgerufen am 17.07.2023).
- Crawford, Kate/Gillespie, Tarleton (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18 (3), S. 410–428.

- DiResta, Renee/Debutts, Matt (2021). "Newsworthiness," Trump, and the Facebook Oversight Board. Hg. v. Columbia Journalism Review. Online verfügbar unter https://www.cjr.org/the_new_gatekeepers/facebook-oversight-board-2.php (abgerufen am 15.07.2023).
- Dolata, Ulrich (2020). Internet – Platforms – Regulation: Coordination of Markets and Curation of Sociality, SOI Discussion Paper, No. 2020-02, Universität Stuttgart, Institut für Sozialwissenschaften, Abteilung für Organisations- und Innovationssoziologie, Stuttgart.
- Douek, Evelyn (2019). Two Calls for Tech Regulation: The French Government Report and the Christchurch Call. Lawfare. Online verfügbar unter <https://www.lawfareblog.com/two-callstech-regulation-french-government-report-and-christchurch-call> (abgerufen am 12.07.2023).
- Facebook (2017). Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism. Facebook Blog. Online verfügbar unter <https://about.fb.com/news/2017/06/global-internet-forum-to-counter-terrorism/> (abgerufen am 20.07.2023).
- Facebook (2021). Facebook veröffentlicht siebten NetzDG-Transparenzbericht. Facebook Blog. Online verfügbar unter <https://about.fb.com/de/news/2021/07/facebook-veroeffentlicht-siebten-netzdg-transparenzbericht/> (abgerufen am 16.07.2023).
- Fielitz, Maik/Schwarz, Karolin (2020). Hate not found. Das Deplatforming der extremen Rechten und seine Folgen. Hg. v. Institut für Demokratie und Zivilgesellschaft. Jena.
- Fielitz, Maik (2022). Die digitalen Sphären des rechten Terrors. Neue Vermittlungsstrategien von Gewalt. In: Jana Kärgel (Hg.). Terrorismus im 21. Jahrhundert. Perspektiven, Kontroversen, Blinde Flecken. Bonn, Bundeszentrale für Politische Bildung, 166–175.
- Fielitz, Maik (2023). Die Verregelung digitaler Hasskommunikation. Eine Chronologie. Hg. v. NEOVEX-Blog. Online verfügbar unter <http://neovex-projekt.de/die-verregelung-digitaler-hasskommunikation-eine-chronologie/> (20.07.2023).
- Flew, Terry/Martin, Fiona (Hg.) (2022). Digital platform regulation. Global perspectives on internet governance. Cham: Palgrave Macmillan (Palgrave global media policy and business).
- Frenkel, Sheera/Kang, Cecilia (2021). Inside Facebook. Die hässliche Wahrheit. Frankfurt am Main: S. Fischer.
- Geschke, Daniel/Klaßen, Anja/ Quent, Matthias/ Richter, Christoph (2019). #Hass im Netz. Der schleichende Angriff auf unsere Demokratie. Hg. v. Institut für Demokratie und Zivilgesellschaft. Jena.
- Gillespie, Tarleton (2018). Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media. New Haven, Yale University Press.
- Gillespie, Tarleton (2022). Do Not Recommend? Reduction as a Form of Content Moderation. Social Media + Society 8 (3).
- Gillespie, Tarleton (2023). The Fact of Content Moderation; Or, Let's Not Solve the Platforms' Problems for Them. MaC 11 (2), 1–4. DOI: 10.17645/mac.v11i2.6610.
- Gorwa, Robert (2019a). What is Platform Governance? Information, Communication & Society 22 (6), 854-871.
- Gorwa, Robert (2019b). The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content. Internet Policy Review 8 (2).
- Gorwa, Robert (2021): Elections, institutions, and the regulatory politics of platform governance: The case of the German NetzDG. Telecommunications Policy 45 (6), DOI: 10.1016/j.telpol.2021.102145.
- Gray, Mary L./Suri, Siddharth (2019). Ghost work. How to stop Silicon Valley from building a new global underclass. Boston, Houghton Mifflin Harcourt.
- Harris, Brent (2020). Preparing the Way forward for Facebook's Oversight Board. Facebook Blog. Online verfügbar unter <https://about.fb.com/news/2020/01/facebooks-oversight-board/> (abgerufen am 14.07.2023).
- Jaspert, Marcel (2023). Are you making Trump great again? Trump darf fortan auch wieder auf Facebook und Instagram hetzen. NEOVEX Blog. Online verfügbar unter <https://neovex-projekt.de/are-you-making-trump-great-again/> (abgerufen am 14.07.2023).
- Kessel, Jeremy (2012). Twitter Transparency Report. Twitter Blog. Online verfügbar unter https://blog.twitter.com/en_us/a/2012/twitter-transparency-report (abgerufen 18.07.2023).

- Kettemann, Matthias C./Fertmann, Martin (2021). Demokratie plattformfest machen. Social Media Councils als Werkzeug zur gesellschaftlichen Rückbindung der privaten Ordnungen digitaler Plattformen. Hg. v. Friedrich Naumann Stiftung, Potsdam.
- Macklin, Graham (2019). The Christchurch Attacks: Livestream Terror in the Viral Video Age. *CTC Sentinel* 12 (6), 18–29.
- Maddox, Jessica/Malson, Jennifer (2020). Guidelines Without Lines, Communities Without Borders: The Marketplace of Ideas and Digital Manifest Destiny in Social Media Platform Policies. *Social Media + Society* 6 (2).
- Magalhães, João Carlos/Katzenbach, Christian (2020): Coronavirus and the frailness of platform governance. *Internet Policy Review*, 9. Online verfügbar unter <https://www.ssoar.info/ssoar/handle/document/68143> (abgerufen am 13.07.2023).
- Marantz, Andrew (2020). Why Facebook cant't fix itself. Hg. v. The New Yorker. Online verfügbar unter <https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself> (abgerufen am 13.07.2023).
- Mau, Steffen (2017). Das metrische Wir. Über die Quantifizierung des Sozialen. 2. Auflage. Berlin, Suhrkamp.
- Noble, Safiya Umoja (2018). Algorithms of oppression. How search engines reinforce racism. New York, New York University Press.
- Pasquale, Frank (2015). The black box society. The secret algorithms that control money and information. Cambridge, Mass., Harvard University Press.
- Roberts, Sarah T. (2019). Behind the screen. Content moderation in the shadows of social media. New Haven, Yale University Press.
- Savolainen, Laura (2022). The shadow banning controversy: perceived governance and algorithmic folklore. *Media, Culture & Society* 44 (6), 1091-1109.
- Seemann, Michael (2020): Die Macht der Plattformen. Politik in Zeiten der Internet-Giganten. 1. Auflage. Berlin, Ch. Links Verlag.
- Serwer, Adam (2023). Elon Musk's Free-Speech Charade Is Over. Unter Mitarbeit von The Atlantic. Online verfügbar unter <https://www.theatlantic.com/ideas/archive/2023/04/elon-musk-twitter-free-speech-matt-taibbis-substack/673698/> (abgerufen am 18.07.2023).
- Stokel-Walker, Chris (2022). Right-Wing Trolls Are Trying to Break Back Into Twitter. Hg. v. Wired UK. Online verfügbar unter https://www.wired.co.uk/article/elon-musk-twitter-trolls?utm_source=twitter&utm_brand=wired&utm_social-type=owned (abgerufen am 15.07.2023).
- Törnberg, Petter (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences of the United States of America* 119 (42), e2207159119. DOI: 10.1073/pnas.2207159119.
- The YouTube Team (2008). Updated Community Guidelines. YouTube Blog. Online verfügbar unter <https://blog.youtube/news-and-events/updated-community-guidelines/> (abgerufen am 14.07.2023).
- The YouTube Team (2020). Safer Internet Day: 5 ways we're building a safer YouTube. YouTube Blog. Online verfügbar unter <https://blog.youtube/news-and-events/safer-internet-day/> (abgerufen am 15.07.2023).
- Twitter (2012). Shutting down spammers. Twitter Blog. Online: https://blog.twitter.com/en_us/a/2012/shutting-down-spammers (abgerufen am 15.07.2023).
- Wojcicki, Susan (2017). Expanding our work against abuse of our platform. YouTube Blog. Online verfügbar unter <https://blog.youtube/news-and-events/expanding-our-work-against-abuse-of-our/> (abgerufen am 15.07.2023).
- Wojcicki, Susan (2020). Mein Halbjahres-Update für die YouTube-Community. YouTube Blog. Online verfügbar unter <https://blog.youtube/intl/de-de/inside-youtube/mein-halbjahres-update-fur-die-youtube-community/> (abgerufen am 16.07.2023).
- York, Jillian C. (2012). YouTube's Censorship of Anti-Islam Video Raises Numerous Questions Around Online Free Expression. Online verfügbar unter <https://www.ef.org/deelinks/2012/09/youtubes-censorship-anti-islam-video-raises-numerous-questions-around-online-free> (abgerufen am 14.07.2023).