DOI: 10.2478/wd-2025-0046

sciendo

Wirtschaftsdienst, 2025, 105(3), 161-166

JEL: H41, L82, Z18

Raphaela Andres, Lara Marie Berger

Digitale Medienmärkte: Was tun gegen Hassrede und Falschinformationen?

Die Digitalisierung verändert die Medienmärkte radikal. Informationen verbreiten sich rasant und Algorithmen entscheiden häufig automatisiert, welche Nachrichten ihren Nutzer:innen angezeigt werden und welche nicht. So sind viele Informationen einer sehr breiten Öffentlichkeit so schnell wie noch nie zugänglich und Menschen weltweit können sich in Echtzeit vernetzen. Was aber auch mit der Digitalisierung von Medienmärkten einhergeht, ist ein stärkeres Aufkommen von Hassrede, Falschinformationen und übermäßig dramatischen Inhalten. Gleichzeitig sind journalistische Qualitätskontrollen wie z. B. Faktenchecker:innen vermehrt Sparmaßnahmen ausgesetzt oder werden – wie kürzlich auf den US-Versionen der Meta-Plattformen Facebook und Instagram – komplett gestrichen.

Um als Gesellschaft angemessen auf diese Veränderungen zu reagieren, ist ein tiefes Verständnis der Funktionsweise digitaler Medienmärkte essenziell. Wie genau sehen Inhalte in einem Umfeld aus, das sich rasant wandelt und in dem traditionelle Erlösmodelle zusehends brüchig werden? Welche Strategien sind hilfreich gegen zirkulierende Falschinformationen, zunehmenden Sensationalismus und Hassrede? Und was kann die Politik tun, damit die Menschen möglichst viel Nutzen aus digitalen Medienmärkten schöpfen? In diesem Beitrag zum Zeitgespräch fassen wir den Stand der Forschung zu diesen Fragen zusammen.

Zunehmende Emotionalisierung von Inhalten

Die moderne Nachrichtenwelt ist von einem starken Wettbewerb um Aufmerksamkeit geprägt, der durch die Digitalisierung weiter verstärkt wird. Ein Beispiel: Wo Zeitungen früher rein aus Platzgründen nur eine Handvoll Schlagzeilen für ihre Titelseite auswählen konnten, publizieren Online-Portale heute dutzende Artikel gleichzeitig, teils mit individuell zugeschnittenen Titeln. Somit konkurriert jede einzelne Überschrift um Klicks (und dadurch um Werbeeinnahmen und Abonnent:innen) und nicht nur die Schlagzeilen auf den Titelblättern.

Zahlreiche Studien zeigen, dass Nutzer:innen besonders stark auf emotionale Inhalte reagieren, insbesondere wenn diese negative Elemente enthalten (Trussler & Soroka, 2014; Soroka, 2016; Soroka et al., 2018). Medienschaffende greifen zudem auf Echtzeitdaten zu, um Nutzer:innenpräferenzen besser zu verstehen. Sie können beispielsweise Klickraten, das Teilen in sozialen Medien sowie die Rezeption in der Kommentarspalte in Echtzeit verfolgen. In der Praxis bedeutet das oft, dass Artikel mit reißerischen Titeln bevorzugt platziert werden, weil sie oftmals für höhere Klickraten sorgen (Leung & Strumpf, 2023; Berger, 2025). Werbeeinnahmen, Paywalls und Abomodelle hängen zunehmend von der Interaktion mit Online-Inhalten ab, sodass der Druck steigt, jedes Thema möglichst gefühlsintensiv zu präsentieren.

Untersuchungen deuten darauf hin, dass eine höhere Emotionalisierung von Schlagzeilen nicht nur mehr Klicks generiert, sondern auch das Leseverhalten und die Wissensaufnahme beeinflusst (Dertwinkel-Kalt et al., 2022; Soroka et al., 2019; Balbuzanov & Tjernström, 2019; Berger, 2025). Gerade die Tendenz, überwiegend negative oder alarmierende Schlagzeilen zu wählen, hat zweischneidige Effekte auf das Publikum. So kann sie einerseits zu längeren Interaktionszeiten mit Nachrichten führen, was aus gesellschaftlicher Sicht im Sinne einer informierten Öffentlichkeit wäre. Gleichzeitig gibt es aber Evidenz für die Entstehung von verzerrten Wahrnehmungen der Wirklichkeit, die wiederum politische Meinungsbildungsprozesse und ökonomische Entscheidungen negativ beeinflussen können (Kayser & Peress, 2021; Berger, 2025).

Dr. Raphaela Andres ist Wissenschaftlerin am ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung in Mannheim.

Lara Marie Berger ist Doktorandin der Volkswirtschaftslehre an der Universität zu Köln.

© Der/die Autor:in 2025. Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht (creativecommons.org/licenses/by/4.0/deed.de).

Open Access wird durch die ZBW – Leibniz-Informationszentrum Wirtschaft gefördert.

Hassrede ist allgegenwärtig und steckt an

Die Zunahme von offen kommuniziertem Hass in den sozialen Medien ist beispiellos. Online-Plattformen werden oftmals als „rechtsfreier Raum“ wahrgenommen, da sich die realen Personen hinter Nutzernamen verstecken und nicht immer ausfindig gemacht werden können. Dies hat zur Folge, dass allein im ersten Quartal 2023 mehr als jede/r vierte Internetnutzer:in Hassreden online wahrgenommen hat (Statistisches Bundesamt, o. D.).

Die Verbreitung von Hassreden im Internet ist alarmierend. Viele Studien belegen die negativen persönlichen, gesellschaftlichen und wirtschaftlichen Konsequenzen. Diese Konsequenzen bleiben nicht nur online, sondern wirken auch in der realen Welt. So verschlechtert sich beispielsweise das individuelle Wohlbefinden der Betroffenen (Braghieri et al., 2022), und das politische Engagement und der soziale Zusammenhalt einer Gesellschaft werden geschwächt (Enikolopov et al., 2020; Bursztyn et al., 2019). Außerdem wurde nachgewiesen, dass Online-Hassreden zu mehr Gewalt in der realen Welt führen (Müller & Schwarz, 2021).

Brandbeschleuniger für Falschinformationen

Die Digitalisierung kurbelt zudem die Verbreitung von faktisch falschen Informationen stark an. Grundsätzlich ist die Verbreitung sogenannter „Fake News“ zwar so alt wie die menschliche Kommunikation. So zirkulierten Unwahrheiten, sobald Informationen weitergegeben werden konnten. Schon 1247 v. Chr. hat etwa Pharao Ramses II. eine verheerende Niederlage in einem Relief als glorreichen Sieg darstellen lassen. Allerdings beschleunigt der technologische Fortschritt die Verbreitung von Informationen und somit auch von Falschinformationen.

In der Folge sind Fake News bei fast allen gesellschaftlichen Themen unserer Zeit präsent: Sei es beispielsweise die erneute Wahl Donald Trumps zum US-Präsidenten, die Coronapandemie oder der russische Angriffskrieg in der Ukraine. Untersuchungen zeigen, dass rund die Hälfte der Nutzer:innen, die Fake News in sozialen Medien sehen, diesen falschen Informationen nach eigenem Bekunden glauben (Allcott & Gentzkow, 2017).

Die Absicht hinter der Verbreitung von Falschnachrichten ist oft politisch: Einige Akteur:innen wollen die öffentliche Meinung manipulieren. Doch auch finanzielle Motive können eine wichtige Rolle spielen. Falschinformationen sind aufgrund fehlender Recherche sehr kostengünstig zu produzieren und generieren viele Klicks, sodass die Erstellung von Falschinformationen auch zum Geschäftsmodell werden kann.

Was also tun gegen Hassrede, Fake News und zunehmende Emotionalisierung? In der Forschung werden unterschiedliche Gegenmaßnahmen diskutiert, die sich grob in folgende Teilbereiche untergliedern lassen:

1. Gegenrede und Faktenchecks
2. Inhaltsmoderation
3. Ökonomische Bedingungen
4. Bildungsmaßnahmen.

Gegenrede: Auf problematische Inhalte reagieren

Ein mögliches Instrument zur Bekämpfung von unerwünschten Online-Inhalten ist die Gegenrede. Als Gegenrede bezeichnet man eine direkte Reaktion auf gefährliche und hasserfüllte Inhalte, um diese zu widerlegen oder zu untergraben (Buerger, 2021).

Wissenschaftliche Belege für die Wirksamkeit von Gegenrede sind bisweilen zumeist deskriptiv oder von geringem Umfang, deuten aber auf die Wirksamkeit dieses Instruments hin. Einen kausalen Nachweis liefern Hangartner et al. (2021), die feststellen, dass Gegenrede, die Empathie hervorruft, die Zahl der nachfolgenden Hassbeiträge reduzieren kann.

Im Kontext von Falschinformationen wird Gegenrede häufig in Form von Faktenchecks, also inhaltlichen Überprüfungen faktischer Behauptungen, bereitgestellt. Professionelle Journalist:innen durchsuchen dafür fortlaufend das Internet nach fragwürdigen Nachrichten und überprüfen diese auf ihre Richtigkeit. Wird eine Nachricht als Falschinformation entlarvt, wird sie auf den Internetseiten der verschiedenen Faktencheck-Organisationen veröffentlicht und korrigiert. Teilweise kooperieren soziale Medien mit den Faktenchecker:innen und ermöglichen es, Falschinformationen als solche auf den Plattformen zu markieren und auf die entsprechende Recherche zu verweisen.

Aus Sicht der Forschung ist allerdings bisher nicht abschließend belegt, ob Faktenchecks überhaupt die gewünschte Wirkung erzielen. Während beispielsweise Bode und Vraga (2015), Vraga und Bode (2017) sowie Henry et al. (2022) deren Wirksamkeit belegen, kommen andere Studien zu dem Ergebnis, dass Faktenchecks entweder gar keine oder unter Umständen sogar sogenannte „Backfire“-Effekte erzeugen können, bei denen es zu einem stärkeren Glauben an die falschen Informationen aufgrund der Faktenchecks kommt (z. B. Nyhan & Reifler, 2010; Nyhan & Reifler, 2015).

Die Wirksamkeit von Faktenchecks variiert zudem in Abhängigkeit vom betrachteten Ziel. Ein Großteil der Litera-

tur ist sich dahingehend einig, dass Faktenchecks zwar nur begrenzte Auswirkungen auf grundlegende Überzeugungen haben, dafür aber effektiv sind, wenn es um konkrete Handlungen geht – beispielsweise die verminderte Weiterverbreitung falscher Nachrichten (Barrera et al., 2020; Nyhan et al., 2020).

Gegenrede ist also eine sinnvolle Maßnahme gegen Hassrede und Falschinformationen, an der sich theoretisch jeder beteiligen und damit die Sicherheit digitaler Medien erhöhen kann. Gleichzeitig hat sie wichtige Einschränkungen. Beispielsweise kann problematischen Inhalten erst eine gewisse Zeit nach ihrer Veröffentlichung widersprochen werden. Sie wurden also vermutlich schon von vielen Nutzer:innen gelesen, bevor sie geprüft und gegebenenfalls markiert werden. Zudem bevorzugen viele Algorithmen neue Inhalte gegenüber älteren. Somit ist ungewiss, ob diese frühen Leser:innen auch die nachträgliche Gegenrede wahrnehmen.

Inhaltsmoderation: Freiwillige, plattforminterne Regeln

Als Reaktion auf die schnelle Verbreitung von Hass und Falschnachrichten im Internet haben viele Plattformen Gemeinschaftsstandards und Hausregeln eingeführt. Häufig weisen Plattformen darauf hin, dass beispielsweise Beleidigungen gegen ihre Hausregeln verstoßen und sie sich somit vorhalten, entsprechende Inhalte löschen zu können.

Allerdings sind die Anreize zur Inhaltsmoderation für Plattformen begrenzt. Einerseits bedarf die Inhaltsmoderation kostenintensiven Einsatz von Technologien und Personal. Außerdem belegen Studien, dass Online-Inhalte, die emotionaler und toxischer sind, häufiger geteilt und kommentiert werden und so zu größeren Netzwerkeffekten für die Plattform führen (Mallipeddi et al., 2021; Beknazar-Yuzbashev et al., 2025). Es ist also nicht unbedingt im Interesse der Plattformen, diese Art von Inhalten zu unterdrücken.

Gleichzeitig hat das Ausmaß an betrügerischem und unsozialem Verhalten bei den Betreibenden von Plattformen zu der Einsicht geführt, dass auch im Internet gewisse Regeln gelten sollten. Diese Einsichten müssen nicht ausschließlich den kommerziellen Interessen der Plattformanbietenden widersprechen, denn unzählige Fake-Accounts, die massenhafte Hetzkommentare und Unwahrheiten verbreiten, gefährden die Sicherheit der Nutzer:innen sowie den Ruf der Werbetreibenden, die nicht mit Hass und Hetze in Verbindung gebracht werden möchten. Der potenzielle Rückzug von Nutzer:innen und Werbetreibenden aufgrund problematischer Inhalte kann also auch das Geschäftsmodell gefährden. So drohten beispielsweise große Marken wie Coca-Cola und Walmart der Videoplattform YouTube, ihre Werbeaktivitäten dort einzustellen, wenn die Plattform ihre Marken nicht besser vor Falschaussagen und extremen Inhalten schützt (Andres et al., 2023).

Am aktuellen Beispiel der USA zeigt sich jedoch, dass freiwillige Selbstregulation von Plattformen stark vom politischen Kontext getrieben sein und sich schnell ändern kann. So entschied Mark Zuckerberg kürzlich, die Zusammenarbeit mit unabhängigen Faktenchecker:innen auf den US-Versionen der Meta-Plattformen wie Facebook und Instagram einzustellen. X (ehemals Twitter) hat eine ähnliche Kooperation bereits kurz nach der Übernahme durch Elon Musk beendet und setzt nun mit sogenannten Community Notes auf Kommentator:innen aus der X-Nutzerschaft statt auf von unabhängigen NGOs finanzierte Faktenchecker:innen.

Inhaltsmoderation: Bindende Regulierung durch Gesetzgeber

Da trotz der oben genannten freiwilligen Maßnahmen viel Hass, Hetze und Falschinformationen verbreitet werden, werden Online-Inhalte auch gesetzlich reguliert. Deutschland hat hier mit dem Netzwerkdurchsetzungsgesetz (NetzDG) eines der weltweit ersten Gesetze zur Eindämmung von Online-Hassrede erlassen, an welchem sich der heutige European Digital Services Act (DSA) stark orientiert. Kerninhalt der Regulierung ist, dass Plattformen ab einer bestimmten Größe für die Inhalte, die sie verbreiten, verantwortlich gemacht werden. Die Plattformen sind verpflichtet, Meldemechanismen einzuführen, durch die Nutzer:innen unerwünschte Inhalte melden können. Die gemeldeten Inhalte müssen zeitnah überprüft und gegebenenfalls von der Plattform gelöscht werden.

Anhand des NetzDGs zeigen Studien, dass Regulierung die Verbreitung von Online-Hassrede eindämmen kann (Jiménez-Durán et al., 2024; Andres & Slivko, 2021). Nach Einführung des NetzDGs wurden politisch sensible Inhalte weniger hasserfüllt und auch die Zahl an Beiträgen mit Hassrede sank. Hierbei wirken drei verschiedene Mechanismen: *Erstens* senkt die Löschung die Zahl an hasserfüllten Beiträgen. Das ist auch dahingehend ein wertvoller mechanischer Effekt, da Studien zeigen, dass toxische Inhalte ansteckend sind, und dass Nutzer:innen, die einer geringeren Toxizität ausgesetzt sind, ihre eigene Toxizität in Beiträgen und Kommentaren reduzieren (Beknazar-Yuzbashev et al., 2025; Thomas & Wahedi, 2023). *Zweitens* passen sich die Nutzer:innen auch an die Regulierung an und posten weniger bzw. weniger hasserfüllte Beiträge, um regelkonform zu agieren. *Drittens* wenden

sich die extrem hasserfüllten Nutzer:innen von den größeren Plattformen ab, um auf kleineren Plattformen, welche nicht unter die Regulierung fallen, weiter zu hetzen. Auch diese Plattform-Migration kann jedoch als Erfolg gewertet werden, da die Reichweite der kleineren Plattformen geringer ist (Rauchfleisch & Kaiser, 2021).

Die belegten Effektgrößen zur Reduzierung von Online-Hassrede sind allerdings gering. Ein Blick in die sozialen Medien genügt, um weiterhin Hass und Hetze ausfindig zu machen. Das ist auch deswegen der Fall, da die Regulierung von Inhalten die Redefreiheit nicht aushebeln darf und die Einordnung einer Aussage als Hassrede schwierig sein kann. Um zu verhindern, dass Plattformen zu viele Inhalte löschen, um Strafen durch die Regulierung zu umgehen, werden nur systematisches Versagen der Plattformen und keine Einzelfallentscheidungen geahndet.

Dennoch belegen die Studien des NetzDGs, dass gesetzlich vorgegebene Inhaltsmoderation zusätzliche Effekte zur freiwilligen Inhaltsmoderation hervorbringen kann. Obwohl X bereits freiwillige Maßnahmen wie Inhaltsmoderation durchführte, wurden die Diskussionen auf X nach Einführung des NetzDGs weniger toxisch (Andres & Slivko, 2021). Darüber hinaus werden Gesetze der gesellschaftlichen Bedeutung digitaler Medien gerecht – wie aktuelle Vorkommnisse verdeutlichen, können freiwillige Plattformmaßnahmen schnell rückgängig gemacht werden, während eine gesetzliche Regelung ein einheitliches und längerfristiges Regelwerk für verschiedene Plattformen festschreibt, die in unserer Gesellschaft gelten sollen. Zudem sorgt allein die Existenz eines Gesetzes dafür, dass digitale Medien weniger als „rechtsfreier Raum“ wahrgenommen werden und sich die Sprache der Nutzer:innen etwas mäßigt (Andres & Slivko, 2021).

Ökonomische Bedingungen

Die Herausforderung für eine neue Medienökonomik besteht darin, tragfähige Geschäftsmodelle zu entwickeln, die weder die Qualität noch die Vielfalt des Informationsangebots gefährden. Neben einer direkten Regulierung digitaler Medienmärkte besteht also die Möglichkeit, ökonomische Anreize so zu setzen, dass die beschriebenen Probleme weniger stark auftreten – beispielsweise durch Subventionierung. Es weisen einige Forschungsarbeiten darauf hin, dass vollständig durch Werbung gestützte und damit von Klicks abhängige Geschäftsmodelle zur weiteren Emotionalisierung beitragen, während alternative Ansätze wie gemeinnütziger Journalismus, Stiftungsfinanzierung, öffentlich-rechtlicher Rundfunk oder ausgewogenere Abomodelle die Sensationslust von Journalist:innen abmildern können (Balbuzanov et al., 2019; Berger, 2025).

Inwiefern eine Veränderung der ökonomischen Anreize die Verbreitung von Hassrede und Falschinformationen einschränkt, ist in der Forschung weniger gut untersucht. Da diese Phänomene aber wie die zunehmende Sensationalisierung durch Aufmerksamkeit der Konsument:innen getrieben werden, ist eine gewisse Abmilderung durch alternativ finanzierte Geschäftsmodelle zumindest plausibel.

Bildungsmaßnahmen

Zusätzlich kann die Medienkompetenz der Nutzer:innen gestärkt werden, sodass sie selbst in der Lage sind, problematische Inhalte zu erkennen und bestenfalls nicht mehr so stark auf diese reagieren. Bachmann et al. (2021) zeigen, dass die Deutschen auf diesem Feld noch viel lernen können: Bei einer Auswertung des Nationalen Bildungspanels (NEPS) finden die Autor:innen altersübergreifende Defizite bei der digitalen Medienkompetenz.

Zahlreiche Studien belegen, dass Medienkompetenz schon durch relativ kurze Trainings verbessert werden kann (z. B. Roozenbeek et al., 2022; Guess et al., 2020; Pennycook et al., 2021; Berger et al., 2025). Darunter werden oft kurze, didaktisch strukturierte Trainings verstanden, die etwa in Schulen, online als Text oder kurzes Video oder durch NGOs angeboten werden und gezielt die Fähigkeit schulen, Falschinformationen zu erkennen und kritisch zu reflektieren. Darüber hinaus unterstützen Untersuchungen von Lewandowsky und Van Der Linden (2021) sowie Ecker et al. (2022) die These, dass ein präventives Vermitteln von Medienkompetenz – oft auch Prebunking oder Inoculation genannt – besonders wirksam sein kann. Das liegt unter anderem daran, dass Nutzer:innen dadurch eine gewisse Resistenz gegenüber später auftretenden Falschinformationen entwickeln, anstatt im Nachhinein versuchen zu müssen, bereits entstandene Fehlvorstellungen zu korrigieren.

Bildungsmaßnahmen wie z.B. Medienkompetenz-Trainings haben gegenüber den bereits besprochenen Gegenmaßnahmen den klaren Vorteil, dass dafür weder die Kooperation der Plattformen noch Abwägungen hinsichtlich Meinungs- und Redefreiheit von Nöten sind. Sie setzen auf nachhaltiges Lernen und Reflektieren, sind einfach umzusetzen und lassen sich kostengünstig auf breite Nutzerkreise anwenden.

Fazit

Digitalisierte Medien können ein großer Gewinn für eine Gesellschaft sein, wenn Phänomene wie Hassrede und Falschinformationen erfolgreich eingegrenzt werden kön-

nen. Eine Schwierigkeit dabei ist, dass schon die Einteilung in problematische Inhalte wie Hassrede und Fake News oft nicht eindeutig und gegebenenfalls sehr aufwendig ist. Beispielsweise spielen im Fall von Hassrede viele gesellschaftliche und zeitliche Komponenten eine Rolle bei der Frage, was gesagt werden darf und was nicht. Bei Fake News kann das Widerlegen falscher Aussagen zeitaufwendig oder teilweise unmöglich sein.

Wie oben beschrieben gibt es aber Gegenmaßnahmen, die die Sicherheit digitaler Medien erhöhen können. Sowohl Gegenrede, Inhaltsmoderation, Veränderungen ökonomischer Bedingungen und Bildungsmaßnahmen werden in der Forschung als vielversprechende Mittel diskutiert. Diese Maßnahmen schließen sich nicht aus und können parallel verfolgt werden.

Wir begrüßen daher die Etablierung der Hassrede-Regelungen im Rahmen des DSA, welche die Problematik auf europäischer Ebene angeht. Einerseits belegen Andres und Slivko (2021), dass eine Regulierung nach den Prinzipien des NetzDGs in der Lage ist, die Verbreitung von Online-Hassrede einzudämmen. Gesetze spiegeln außerdem die gesellschaftliche Relevanz digitaler Medien wider: Während Plattformen ihre Maßnahmen rasch widerrufen können, sorgt eine gesetzliche Regulierung für ein einheitliches, langfristiges Regelwerk. Allein das Bestehen einer gesetzlichen Grundlage signalisiert zudem, dass der digitale Raum kein rechtsfreier Raum ist.

Allerdings sind die bisherigen Regelungen sehr träge. Obwohl beispielsweise mehrere laufende Verfahren der EU Kommission gegen X geführt werden, sind Hass und Hetze auf X noch allgegenwärtig. Um dies zu beschleunigen, könnte man das System der „Trusted Flaggers“ ausbauen. Diese sind im Rahmen des DSAs ernannte Stellen, die mit dem Auffinden illegaler Online-Inhalte betraut sind und deren Meldungen von den Plattformen priorisiert geprüft werden müssen. Bislang ist die Zahl an Trusted Flaggers gering, auch aufgrund der hohen Auflagen, die diese zu erfüllen haben. Da die Trusted Flaggers eine hohe Treffsicherheit in Bezug auf illegale Inhalte haben, könnte deren Ausbau die Inhaltsmoderation verbessern.

Des Weiteren sprechen wir uns für einen Ausbau der Bildungsmaßnahmen zur Förderung der Medienkompetenz aus. Einerseits besteht hier kein Konflikt zur Wahrung von Rede- und Meinungsfreiheit, wie es bei direkten Eingriffen wie Gegenrede und Inhaltsmoderation der Fall ist. Zudem wirkt Bildung präventiv und langfristig: Nutzer:innen profitieren davon auch, wenn sie problematischen Inhalten begegnen, die gerade erst veröffentlicht wurden. Außerdem kann – ähnlich zu den gesetzlichen Regulierungen – Medienkompetenz nicht einfach rückgängig gemacht oder aufgekündigt werden. Dass dies ein wichtiger Faktor ist, verdeutlichen die jüngsten Entwicklungen der amerikanischen sozialen Medien.

Darüber hinaus ist weitere Forschung unerlässlich. Digitalisierte Medien haben einen zu großen Einfluss auf unsere Gesellschaft, als dass die Entscheidungen, was gesagt werden kann und wie, alleine in den Händen profitorientierter Unternehmen gelegt werden kann. Um digitale Medienmärkte so weiterzuentwickeln, dass sie möglichst sicher für ihre Nutzer:innen sind, ist es also essentiell, dass die Forschung einen umfangreicheren Zugang zu Plattformdaten erhält. Hier besteht dringender Handlungsbedarf: Zwar ist der Datenzugang für die Forschung über den DSA festgeschrieben, allerdings ist dieser in der praktischen Umsetzung nur sehr schwierig und zeitverzögert zu erhalten.

Literatur

Allcott, H. & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Andres, R., Rossi, M. & Tremblay, M. (2023). YouTube “Adpocalypse: The YouTubers’ Journey from Ad-Based to Patron-Based Revenues. *ZEW – Leibniz Centre for European Economic Research Discussion Paper*, Nr. 23-059.

Andres, R. & Slivko, O. (2021). Content Regulation or Self-Moderation? The Effect of the Network Enforcement Act on Twitter. *ZEW – Leibniz Centre for European Economic Research Discussion Paper,* Nr. 21-103.

Bachmann, R., Hertweck, F., Kamb, R., Lehnert, J., Niederstadt, M. & Rulf, C. (2021) Digitale Kompetenzen in Deutschland – eine Bestandsaufnahme. *RWI Materialien*, Nr. 150.

Balbuzanov, I., Gars, J. & Tjernström, E. (2019). *Media and motivation: The effect of performance pay on writers and content* (Working Paper).

Barrera, O., Guriev, S., Henry, E. & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, *182*, 104123.

Beknazar-Yuzbashev, G., Jiménez-Durán, R., McCrosky, J. & Stalinski, M. (2025). Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment. *CESifo Working Paper*, Nr. 11644.

Berger, L. (2025). *How digital media markets amplify news sentiment* (Working Paper).

Berger, L., Kerkhof, A., Mindl, F. & Münster, J. (2025). Debunking “fake news” on social media: Immediate and short-term effects of fact-checking and media literacy interventions. *Journal of Public Economics*, Forthcoming.

Bode, L. & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, *65*(4), 619–638.

Braghieri, L., Levy, R. & Makarin, A. (2022). Social media and mental health. *American Economic Review*, *112*(11), 3660–3693.

Buerger, C. (2021, 01. Juni). Counterspeech: A Literature Review. *Dangerous Speech Project*.

Bursztyn, L., Egorov, G., Enikolopov, R. & Petrova, M. (2019). *Social media and xenophobia: evidence from Russia*. National Bureau of Economic Research.

Dertwinkel-Kalt, M., Münster, J. & Zegners, D. (2022). *If it Bleeds, it Leads: Negativity in Online New* (Working Paper).

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K. & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29.

Enikolopov, R., Makarin, A. & Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, *88*(4), 1479–1514.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J. & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27).

Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demircia, B., Derksen, L., Hall, A., Jochum, M., Murias Munoz, M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardic, F. & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, *118*(50).

Henry, E., Zhuravskaya, E. & Guriev, S. (2022). Checking and Sharing Alt-Facts. *American Economic Journal: Economic Policy*, *14*(3), 55–86.

Jiménez-Durán, R., Müller, K. & Schwarz, C. (2024). *The effect of content moderation on online and offline hate: Evidence from Germany's NetzDG*.

Leung, T. C. & Strumpf, K. S. (2023). *All the Headlines that Are Fit to Change*.

Lewandowsky, S. & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348–384.

Mallipeddi, R., Janakiraman, R., Kumar, S. & Gupta, S. (2021). The effects of social media content created by human brands on engagement: Evidence from Indian general election 2014. *Information Systems Research*, *32*(1), 212–237.

Müller, K. & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, *19*(4), 2131–2167.

Nyhan, B., Porter, E., Reifler, J. & Wood, T. J. (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, *42*(3), 939–960.

Nyhan, B. & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330.

Nyhan, B. & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*(3), 459–464.

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D. & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595.

Rauchfleisch, A. & Kaiser, J. (2021). *Deplatforming the far-right: An analysis of YouTube and BitChute*.

Roozenbeek, J., Van der Linden, S., Goldberg, B., Rathje, S. & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34).

Kayser, M. A. & Peress, M. (2021). Does the Media Cover the Economy Accurately? An Analysis of Sixteen Developed Democracies. *Quarterly Journal of Political Science*, *16*(1), 1–33.

Soroka, S. (2016). Gatekeeping and the Negativity Bias. *Oxford Research Encyclopedia of Politics*.

Soroka, S., Daku, M., Hiaeshutter-Rice, D., Guggenheim, L. & Pasek, J. (2018). Negativity and Positivity Biases in Economic News Coverage: Traditional Versus Social Media. *Communication Research*, *45*(7), 1078–1098.

Soroka, S., Fournier, P. & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, *116*(38), 18888–18892.

Statistisches Bundesamt. (o. D.). *Gut ein Viertel der Internetnutzenden stößt im Netz auf „Hatespeech"*.

Thomas, D. R. & Wahedi, L. A. (2023). Disrupting hate: The effect of deplatforming hate organizations on their online audience. *Proceedings of the National Academy of Sciences*, *120*(24).

Trussler, M., and Soroka, S. (2014). Consumer Demand for Cynical and Negative News Frames. *International Journal of Press/Politics, 19*(3), 360–379.

Vraga, E. K., and Bode, L. (2017) Using expert sources to correct health misinformation in social media. *Science Communication, 39*(5), 621–645.

Title: ***Digital Media Markets: How to Curb Hate Speech and Fake News?***

Abstract: *A core challenge for media economics in the digital age is to establish frameworks that neither endanger user safety nor undermine the diversity of information. Hate speech, misinformation and sensationalism must not take over, yet freedom of opinion and expression must remain intact. To bolster digital media security, researchers explore multiple measures. Counter-speech, content moderation, shifts in economic structures, and educational initiatives are discussed as promising strategies. These measures are not mutually exclusive and should be pursued in parallel.*