

Kommunikations-
und Medienethik

Stapf | Heesen | Martena [Hrsg.]

Zwischen Hassrede, Framing und generativer Künstlicher Intelligenz

Medien und Sprache aus
ethischer Perspektive



Nomos

Kommunikations- und Medienethik

Herausgegeben von
Alexander Filipović
Christian Schicha
Ingrid Stapf

Band 24

Ingrid Stapf | Jessica Heesen | Laura Martena [Hrsg.]

Zwischen Hassrede, Framing und generativer Künstlicher Intelligenz

Medien und Sprache aus
ethischer Perspektive



Nomos

Die in diesem Sammelband präsentierten Beiträge durchliefen ein Peer-Review-Verfahren zur Qualitätssicherung und wurden von unabhängigen Fachvertreter:innen begutachtet.

Für die finanzielle Unterstützung, die die Bereitstellung dieses Bandes open access ermöglicht hat, bedanken wir uns bei der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPK), der Universität Greifswald, der Hochschule Mittweida und der Hochschule Bonn-Rhein-Sieg.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2026

© Die Autor:innen

Publiziert von
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-7560-4020-9

ISBN (ePDF): 978-3-7489-7032-3

DOI: <https://doi.org/10.5571/9783748970323>



Onlineversion
Inlibra



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Inhaltsverzeichnis

Laura Martena / Ingrid Stapf / Jessica Heesen

Einleitung 9

I. Umstrittene Sprache: Hate Speech und Gender-Diskurse

Inga Bones

Was ist Hassrede? Versuch einer Schärfung eines umstrittenen Begriffs 21

Bernhard Debatin

„Squatting for Hitler“: Hate Speech, Datenmüll, und generative KI in Zeiten rechtspopulistischer Autokratie 33

Jörg-Uwe Nieland

„Auf geht’s, kämpfen und siegen!“ Fan-Kommunikation zwischen Positionierung, Gegnerschaft und Ausgrenzung 51

Miriam Goetz und Vanessa Wimmers

Wahrnehmung, Einfluss und Folgen sexistischer Hate Speech auf Instagram 73

Beatrice Dernbach

Die Positionen von Redaktionen zur Verwendung gendersensibler Sprache 89

II. Automatisierte Sprache: Potenziale und Risiken Künstlicher Intelligenz

Brigitte Huber und Julia Levasier

„Ich kenne mich mit KI-Technologie nicht aus, aber...“: Eine wissenssoziologische Diskursanalyse von User:innendiskussionen zu Künstlicher Intelligenz auf Nachrichtenwebsite und *Instagram*-Account von *Die Zeit* 111

Michael Litschka

KI-Modelle in Medienunternehmen: empirische Befunde und ethische Reflexionen für die Regulierung 133

Matthias O. Rath

„Artificial Sepsis“ – Leitlinien einer salutogenetischen Bot-Nutzung 151

Jana Hecktor

Ethische Perspektiven auf Große Sprachmodelle am Beispiel von Trainingsdatenqualität 167

Theresa Krampe

Imagining Fair(er) Datasets for GenAI: Lessons from the Arts 187

Mario Anastasiadis und Hektor Haarkötter

Der Algorithmus macht, was er soll, oder? – Eine technikethische Reflexion automatisierter Detektion von Desinformationen im Internet 211

III. Öffentliche Sprache: Ethische Perspektiven auf demokratische Debattenkultur

Petra Grimm, Susanne Kuhnert und Marcel Schlegel

Ausgewählte Perspektiven auf den Einsatz von Künstlicher Intelligenz in den Medien 233

Christian Schicha

Kommunikationsfreiheiten und Öffentlichkeiten als Basis
verständigungsorientierter Streitkulturen in der Demokratie 257

Stefanie Aeverbeck-Lietz

Geltungsansprüche „pluralistisch-transzendentaler Öffentlichkeit“
als Grenzziehungen gegen holistische „qualitative Öffentlichkeit“.
Überlegungen zur Integration der Ansätze von Habermas und
Manheim im Rekurs auf Seyla Benhabib 277

Henning Eichler

Öffentlich-rechtliche Medien als Gestalter konstruktiver
Debattenräume: der Public Spaces Incubator 299

Janis Brinkmann

Qualitätsjournalismen? Eine (Re-)Definition flexibler Kriterien der
Qualität alternativer journalistischer Berichterstattungsmuster 317

Verzeichnis der Autorinnen und Autoren 337

Einleitung

Laura Martena / Ingrid Stapf / Jessica Heesen

Der Ausdruck „Kommunikation“ stammt von „communio“ ab, was „mitteilen“, aber auch „teilen“, „gemeinsam machen“ oder gar „vereinigen“ heißen kann. Kommunikation, im Wesentlichen sprachlich verfasste, ermöglicht einerseits den Ausdruck des Eigenen, andererseits Differenz und Pluralität, und vermag idealiter gerade darin Gemeinsamkeit zu schaffen. Damit wird sie zum wesentlichen Medium moralischer Aushandlungsprozesse und demokratischer Verständigung.

Im demokratischen Diskurs selbst wurde Sprache lange als vermeintlich neutrales Werkzeug behandelt, das Inhalte transportiert, ohne selbst Thema zu werden. In den vergangenen Jahren hat sich dies jedoch gewandelt. Der Sprachgebrauch ist heute selbst zu einem zentralen Gegenstand moralischer und politischer Konflikte geworden. Immer neue Debatten um Formulierungen, die früher als unproblematisch galten, jetzt aber als diskriminierend wahrgenommen werden und deshalb aus dem Wortschatz gestrichen werden sollen, oder auch die Forderungen nach geschlechtergerechter Sprache verweisen darauf. Gleichzeitig versuchen insbesondere rechtsextreme Akteure, oft unter Berufung auf die Redefreiheit, gezielt in den privaten und öffentlichen Sprachgebrauch zu intervenieren. Im Rahmen ihrer „metapolitischen“ Strategien, die sie ursprünglich linken Kontexten entnommen haben, arbeiten sie darauf hin, diskreditierte Begriffe zu rehabilitieren oder neue zu etablieren, um Wahrnehmungsbereitschaften und Bewertungsmaßstäbe im Sinne ihrer eigenen politischen Agenda zu verschieben (vgl. Ruppert-Karakas 2024). Diese gegenläufigen Dynamiken – die Sensibilisierung für diskriminierende und verletzende Sprache einerseits, ihre gezielte ideologische Instrumentalisierung andererseits – lenken den Blick auf neue Weise auf die wirklichkeitskonstituierende Macht der Sprache. Sprache wird in dieser Sicht zum politischen Kampfplatz.

Diese Entwicklungen vollziehen sich in einem medialen Umfeld, das sich im Zuge der Digitalisierung tiefgreifend verändert hat. Soziale Medien, die anfänglich mit hohen Erwartungen an eine Demokratisierung des Diskurses verbunden waren, weil jeder nun zugleich Sender:in und Empfänger:in sein können sollte, sind Arenen der Polarisierung geworden.

Hassrede, Beleidigungen und Diffamierungen prägen die kommunikative Praxis vieler Plattformen. Obwohl Dienste wie Instagram, TikTok oder X durch ihr Content-Management maßgeblich mitbestimmen, welche Inhalte sichtbar werden, verstehen sie sich selbst meist als neutrale Vermittler und weisen Verantwortung von sich. Die bislang ohnehin nicht sehr effektiven Mechanismen der Moderation und Regulierung sind angesichts der schieren Menge an Inhalten prekär geblieben und politisch umstritten: Was die einen als Voraussetzung für faire Diskurse begreifen, sehen andere gerade als Eingriff in die Meinungsfreiheit, der den demokratischen Diskurs untergrabe. In den USA werden Moderationen öffentlichen Sprachgebrauchs unter Donald Trump derzeit zunehmend wieder abgeschafft (vgl. Debatin in diesem Band), während die Europäische Union mit dem *Digital Services Act* die Funktionen einer demokratischen Öffentlichkeit stützen will.

In dieses Spannungsfeld ist mit Generativer Künstlicher Intelligenz eine neue, potenziell disruptive Kraft getreten. Einerseits versprechen Große Sprachmodelle enorme Potenziale für journalistische Arbeit und politische Kommunikation. Im professionellen Medienbereich werden sie bereits jetzt verwendet, um z. B. Wetter-, Börsen- oder Sportmeldungen zu verfassen. Auch in Film, Kunst und Musik ist Generative Künstliche Intelligenz in Produktionen integriert und Teil öffentlich zugänglicher Inhalte, ebenso in Form virtueller Influencer in den Sozialen Medien. Neben den Potenzialen werden seit einigen Jahren auch die Gefahren und Möglichkeiten des Missbrauchs von Künstlicher Intelligenz diskutiert. So neigen Modelle, die synthetische Inhalte generieren, zu sogenannten Halluzinationen, das heißt dazu, überzeugend formulierte Ergebnisse auszugeben, die objektiv gesehen aber falsch sind. Auch wenn dies ohne Absicht geschieht, können diese Misinformationen (vgl. Aïmeur/Amri/Brassard 2023) die Vertrauenswürdigkeit öffentlicher Kommunikation gefährden. Generative Künstliche Intelligenz wird zudem gezielt für Desinformation missbraucht. Diese lässt sich unter Verwendung entsprechender Tools schneller in großer Zahl erzeugen und leichter personalisiert verbreiten. Vor allem im Umfeld von Wahlen und insgesamt im Zusammenhang der öffentlichen und individuellen Meinungsbildung entfalten Deepfakes als Desinformation von besonders hoher Suggestions- und Täuschungskraft zunehmend an Wirkung (vgl. Muñoz 2024). Insbesondere dann, wenn künstlich generierte Mis- oder Desinformationen wiederum durch KI-Anwendungen verwendet werden, kann ein gefährlicher Kreislauf entstehen, der nur noch schwer durch menschliche Aufsicht kontrolliert und korrigiert werden kann.

Angesichts dieser Phänomene lässt sich von einer multiplen Krise demokratischer Debattenkultur sprechen. Sie betrifft die symbolischen Grundlagen politischer Verständigung, die institutionellen Bedingungen öffentlicher Kommunikation und die technischen Infrastrukturen gleichermaßen. Die beschriebene Krise ist zugleich die Stunde der Kommunikations- und Medienethik. Ethik ist immer auch „Krisenreflexion“ (Riedel 1979). In diesem Sinn übt die Kommunikations- und Medienethik Kritik an bestehenden Kommunikationsordnungen und ihren Infrastrukturen, indem sie deren Legitimität, Angemessenheit und demokratische Tragfähigkeit befragt. Zugleich gibt Ethik Orientierung für die Praxis (Debatin 1997). Die Kommunikations- und Medienethik macht konkret normative Orientierungsangebote für diejenigen, die in Medien und Öffentlichkeit handeln – etwa in Gestalt von Kodizes, Qualitätsstandards oder berufsethischen Leitlinien. So kann Sie Maßstäbe geben, an denen sich die Gesellschaft und gesellschaftliche Verständigung in ihrer Diversität ausrichten kann. In der gegenwärtigen Konstellation kann die Kommunikations- und Medienethik einen Beitrag zur Sicherung und Erneuerung demokratischer Öffentlichkeit(en) leisten – und wird damit selbst zu einer wichtigen Stimme in der Auseinandersetzung um die Bedingungen verantwortlicher Kommunikation.

Der vorliegende Band versteht sich als ein solcher Beitrag. Er widmet sich dem Verhältnis von Medien und Sprache aus ethischer Perspektive und versammelt theoretische, normative und empirische Studien. Gemeinsam verfolgen sie das Ziel, sprachbezogene Konflikte, digitale Kommunikationspraktiken und KI-gestützte Informationsprozesse in ihren ethischen Implikationen sichtbar zu machen und Orientierungen für eine demokratische Kommunikationskultur zu entwickeln. Zusammen geben die Beiträge eine medienethische Bestandsaufnahme einer Gegenwart, in der Sprache nicht nur Mittel, sondern Gegenstand öffentlicher Konflikte ist, digitale Infrastrukturen die Kommunikation prägen und neue Technologien die Schaffung geteilter Welten herausfordern.

Teil 1: Die Beiträge sind drei Themenschwerpunkten zugeordnet. Die Autor:innen des ersten Teiles über *umstrittene Sprache* widmen sich Sprache als Medium und Gegenstand ethischer Auseinandersetzungen anhand konkreter Konfliktfelder. Ein erster Schwerpunkt liegt auf Hassrede. Während dieser Begriff im öffentlichen Diskurs in den vergangenen Jahren allgegenwärtig geworden ist, empirische Studien ihre Verbreitung aus Sicht Betroffener nahelegen und Gegenmaßnahmen gefordert werden, wird seltener gefragt, was „Hassrede“ genauer ist. Das führt zu einer uneinheitli-

chen Verwendung, die auch die Einordnung entsprechender Studienergebnisse und die Evaluation von Maßnahmen erschwert. Dieser Frage widmet sich Inga Bones anhand einer kritischen Diskussion bestehender Begriffsbestimmungen. Sie kommt zu dem Schluss, dass Hassrede am besten als Form der Kommunikation zu verstehen ist, die eine Person oder Gruppe auf Basis stabiler und identitätsrelevanter Eigenschaften wie Religion, Ethnie, Geschlecht oder Herkunft auf tiefgreifende Weise angreift, abwertet oder diskriminiert.

Bernhard Debatin zeigt auf, welcher Strategien sich rechtsextreme Akteure in den USA bedienen, um Hassrede zu verbreiten. Besonders beliebt sei das sogenannte *Dog Whistling*. Durch kodierte und suggestive Aussagen werde dabei Hass auf Angehörige bestimmter Gruppen bei der eigenen Klientel weiter geschürt. Außenstehenden erschließe sich deren rassistischer und gewaltverherrlichender Gehalt dagegen nicht unmittelbar, sodass dieser sich im Zweifel leicht verleugnen und der Rassismuskritik gegen diejenigen kehren lasse, die ihn erheben. Dabei bedienen sich die entsprechenden Akteur:innen auch Generativer KI, um ihre Hassbotschaften weiter zu verbreiten.

Jörg-Uwe Nieland untersucht das Verhältnis von Sprache, Emotion und Konflikt in der organisierten Sportfankultur. Durch die Digitalisierung verlagern sich auch hier Auseinandersetzungen rivalisierender Fangruppen in digitale Arenen. Anhand einer Pilotstudie zum österreichischen Eishockey zeichnet der Autor nach, wie Fankommunikation hier stattfindet. Dabei lasse sich unter den Fans eine grundlegend agonistische Online-Kommunikation beobachten, die von Rivalität, Abwertung des Gegners und Aufwertung des eigenen Vereins und seiner Anhänger:innen gekennzeichnet sei. Bei alledem bleibe die Streitkultur aber zivilisiert. Dehumanisierende Feindbilder und Hassrede träten auf den Instagram-Kanälen der Vereine nicht offen zutage. Ob sich die entsprechende Kommunikation in andere Kanäle verlagert hat, bleibt offen.

Miriam Goetz und Vanessa Wimmers beleuchten spezifisch sexistische Hate Speech im Sinne geschlechtsspezifischer Diskriminierung, die vor allem Frauen – hier exemplarisch bei Instagram – betrifft. Userinnen berichteten in hohem Maße von Beleidigungen, sexualisierter Belästigung sowie Vergewaltigungs- und Morddrohungen in Kommentaren und Direktnachrichten. Instagram-spezifische Funktionen wie Direktnachrichten und die niedrige Zugangsschwelle zur Kontaktaufnahme begünstigten gezielte Angriffe. Dabei wirkten algorithmische Verstärkungsmechanismen und gruppendynamische Effekte normalisierend auf sexistische Inhalte. Mel-

de- und Moderationssysteme würden von den Nutzerinnen als ineffektiv wahrgenommen. Entsprechend würden härtere Sanktionen, bessere Filter, menschliche Moderation, geringere Anonymität und rechtliche Durchsetzung gefordert.

Wie gehen Medien mit der Politisierung der Sprache um? Dieser Frage widmet sich Beatrice Dernbach am Beispiel des Genderns. Sie rekonstruiert, wie unterschiedliche deutschsprachige Redaktionen sich im Streit um gendersensible Sprache positionieren – von konsequenter Nutzung des Genderns mit Sonderzeichen über pragmatische, eher zurückhaltende Lösungen bis zu klarer Ablehnung von Sonderzeichen und Betonung des generischen Maskulinums. Die Verfasserin zeigt, welche normativen Vorstellungen der Sprache und der eigenen gesellschaftlichen Rolle dahinterstehen, um schließlich für eine Versachlichung der Debatte zu plädieren. Dabei diskutiert sie auch, inwiefern der wachsende Einsatz Generativer Künstlicher Intelligenz in den Redaktionen zu einer solchen Versachlichung beitragen oder aber die Polarisierung weiter verstärken könne.

Teil 2: Auf die Analyse aktueller Sprachkonflikte folgt ein Blick auf *automatisierte Sprache*, und damit die technologischen Veränderungen, die die Produktion von Sprache selbst transformieren. Der zweite Teil des Bandes befasst sich mit ethischen Fragen, die speziell mit dem Aufstieg Generativer Künstlicher Intelligenz, insbesondere Großen Sprachmodellen verbunden sind. Er beleuchtet diese Problemstellungen auf der Ebene des öffentlichen Diskurses über KI, auf der gesamtgesellschaftlichen Ebene, hinsichtlich ihres Einsatzes in Redaktionen und auf der Mikroebene der Trainingsdaten. Und er fragt, wie KI-Anwendungen ihrerseits helfen können, einige dieser Probleme zu lösen.

Zunächst werfen Brigitte Huber und Julia Levasier Schlaglichter auf aktuelle Online-Diskussionen über KI im deutschsprachigen Raum seit der Veröffentlichung von ChatGPT. Anhand einer Untersuchung der Diskurse auf der Nachrichtenwebsite und dem Instagram-Account von *Die Zeit* zeigen sie, dass diese von ethischen Bedenken und Szenarien des Kontrollverlusts geprägt sind. Dazu zählt etwa die der Science Fiction entstammende Vorstellung, KI könne ein Bewusstsein entwickeln, aber auch die handfeste Sorge vor Diskriminierung und mangelndem Datenschutz. Aus wissenssoziologischer Sicht zeichnen sie nach, dass die Nutzer:innen dabei vor allem auf Erfahrungswissen und populärkulturelle Referenzen zurückgriffen, während wissenschaftliches Wissen selten rezipiert werde. Zugleich neigten die Nutzer:innen dazu, die Wissensansprüche anderer in Zweifel zu ziehen.

Michael Litschka stellt die Ergebnisse einer Studie zum aktuellen Einsatz von KI-Anwendungen, insbesondere Sprachmodellen, in österreichischen und internationalen Medienunternehmen aus Sicht der Akteur:innen dar. Da diese Erhebung neben ökonomischen Gesichtspunkten auch ethische Bedenken und ein Bedürfnis nach Regulierung offenbart, geht er im zweiten Schritt der Frage nach Möglichkeiten und Grenzen solcher Regulierungsversuche und ihrer ethischen Fundierung nach. Anhand der Gerechtigkeitstheorien von John Rawls und Amartya Sen liefert er den Befürworter:innen von Regulierung eine medienethische Begründung und erweitert die bestehenden KI-ethischen Diskurse um eine seltener eingenommene Perspektive.

Matthias Rath beleuchtet eine systematische Gefahr Generativer KI, die im öffentlichen Diskurs weniger präsent ist: eine „artifizielle Sepsis“. Mit dieser Metapher beschreibt er, wie der unregulierte Einsatz solcher Tools zu einer epistemischen „Selbstvergiftung“ führen könne. Diese drohe dann, wenn Generative KI-Modelle beginnen, wiederum maschinell erzeugte Inhalte, die schon heute einen Großteil der online verfügbaren Inhalte ausmachen, in ihre Such- und Trainingsprozesse einzuspeisen. Dadurch würde sich die Informationsökologie schrittweise von ihrer Begründung in realweltlicher Erfahrung entkoppeln. Als Gegenmodell schlägt der Autor eine „salutogenetische Bot-Nutzung“ vor, die epistemische Resilienz stärken soll.

Die Gefahr „artifizieller Sepsis“ verweist auf die zentrale Bedeutung der Qualität von Trainingsdaten, der sich die nächsten beiden Beiträge widmen. Wenn Jana Hecktor Qualitätskriterien von Trainingsdaten für KI-Anwendungen im Allgemeinen vorstellt und diese speziell auf LLMs anwendet, bezieht sie nicht nur die Auswahl bestehender Datensätze ein, sondern auch den Prozess der Datengewinnung – einschließlich der Rechte derjenigen Menschen, die bewusst oder unbewusst daran mitarbeiten. Konkret rückt sie Ausgewogenheit, Datenschutz, Diversität, Fairness, Korrektheit, Privatsphäre und Repräsentativität in den Fokus und buchstabiert aus, was damit im Kontext von LLMs verbunden wäre. Dabei könnten die Kriterien im Einzelfall durchaus in Widerspruch geraten und sollten fallbezogen gegeneinander abgewogen werden. Die Anwendung solcher Kriterien könnte helfen, offensichtlicher Diskriminierung durch Generative KI-Tools vorzubeugen.

Gibt es aber nicht auch subtilere, verdecktere und daher schwerer detektierbare Formen geschlechtsbezogener, rassistischer oder ableistischer Diskriminierung und Stereotypisierung? Wenn ja, wie ließen diese sich im

Maschinellen Lernen aufdecken und vermeiden? Anregungen findet Theresa Krampe in zeitgenössischen Kunstprojekten, die sich Generativer KI auf kreative und subversive Weise bedienen. Solche Projekte, die mit diversifizierten und sorgfältig kuratierten Datensätzen arbeiten, könnten nicht nur hegemoniale Vorstellungen irritieren, die unserem aktuellen Umgang mit KI oft zugrunde liegen, und *biases* zu minimieren helfen. Sie könnten auch dazu inspirieren, alternative und fairere KI-Zukünfte zu imaginieren.

Können automatisierte Systeme auch helfen, die Probleme zu lösen, die sie selbst mit erzeugen oder verstärken? Dieser Frage widmen sich Mario Anastasiadis und Hektor Haarkötter am Beispiel von Desinformation. Deren enorme Verbreitung im Netz droht die öffentliche Meinungsbildung zu verzerren, Diskursräume zu destabilisieren und kann das Vertrauen in demokratische Institutionen beschädigen. Um dagegen anzugehen, wurde die App NEBULA entwickelt, die Desinformation automatisch detektieren soll. Die Autoren stellen Ergebnisse der medienethischen Begleitforschung zur Entwicklung dieser App vor. NEBULA soll nicht nur Hinweise auf Desinformation ausgeben, sondern die Fähigkeit stärken, diese eigenständig zu erkennen.

Teil 3: Die Beiträge des dritten Teils über *öffentliche Sprache* weiten den Blick, indem sie die Frage nach den normativen Grundlagen demokratischer Streitkultur in Zeiten der Digitalisierung und des Aufstiegs Generativer Künstlicher Intelligenz neu stellen.

Petra Grimm, Susanne Kuhnert und Marcel Schlegel berichten von zwei Forschungs- und Lehrprojekten, die sich um die Auswirkungen Generativer KI auf das Mediensystem drehen. Im Fokus stehen Anwendungen zur artifiziellen Nachbildung menschlicher Stimmen. Ein erstes Projekt richtete sich an Medienstudierende, die anhand neuer didaktischer Szenarien, insbesondere realitätsnaher Fallstudien, ethische Aspekte des Umgangs damit abzuwägen lernen sollten. Ein zweites Projekt nahm erfahrene Medienpraktiker:innen in den Blick. Anhand qualitativer Interviews wurden zunächst Leitlinien zur KI-Nutzung erarbeitet, die Medienschaffenden helfen sollten, die ethische Legitimität eines spezifischen KI-Einsatzes zu überprüfen.

Christian Schicha stellt die Bedeutung der verfassungsrechtlich garantierten Meinungs- und Kommunikationsfreiheit in den Mittelpunkt. Ausgehend von Jürgen Habermas skizziert er ein Konzept deliberativer Öffentlichkeit, zu der alle Bürger:innen gleichberechtigt Zugang haben und in der Argumente statt Machtmittel zählen. Der digitale Strukturwandel drohe dieses Ideal zu untergraben, indem er die Entstehung fragmentierter, vermachteter und algorithmisch gesteuerter Teil- und Gegenöffentlichkeiten fördere. Dagegen

setzt der Autor das Ideal einer an Solidarität, Empathie, Respekt und wechselseitiger Anerkennung orientierten Streitkultur, die allein demokratische Selbstbestimmung und Gemeinwohlorientierung sichere.

Stefanie Averbek-Lietz begründet den Pluralismus als Wert aus kommunikationsethischer Perspektive und verankert ihn in einem Universalismus der Würde. Mit Seyla Benhabib skizziert sie einen „interaktiven Universalismus“, in dem moralischer Universalismus und die Anerkennung der Andersheit der Anderen sich nicht länger ausschließen, sondern wechselseitig bedingen und hervorbringen. Auch sie knüpft an Jürgen Habermas an, wenn sie diesem Gedanken eine kommunikationstheoretische Wendung gibt. Schließlich unterscheidet sie mit Ernst Manheim verschiedene Konzeptionen der Öffentlichkeit, die dieses Ideal ermöglichen oder untergraben. Der Beitrag skizziert die Aufgabe einer kommunikationsethischen Bildung, die Anerkennung von Verschiedenheit, Sensibilität für verletzte Geltungsansprüche und Widerstandskraft stärkt – auch mit Blick auf KI-gestützte Kommunikationsumgebungen.

Wie können konstruktive Debattenräume im Digitalen geschaffen werden, wenn kommerzielle Plattformen einzig an Gewinnmaximierung durch höhere Verweildauer der Nutzer:innen interessiert sind und dazu polarisierende Inhalte bevorzugen? Was haben die öffentlich-rechtlichen Medien dem entgegenzusetzen? Henning Eichler stellt mit dem *Public Spaces Incubator* (PSI) ein internationales Forschungs- und Entwicklungsprojekt vor, das die Sender wenn nicht ganz aus der Abhängigkeit von kommerziellen Plattformen lösen, so doch neue Räume für wertorientierte, konstruktive und respektvolle Online-Debatten eröffnen könne. Damit wollen sie ihrem Funktionsauftrag auch angesichts veränderter Erwartungen des Publikums an den Dialog mit Journalist:innen wieder gerecht werden. Eichler stellt die Ergebnisse einer qualitativen Studie zu ethischen Herausforderungen dar, die sich aus Sicht von Journalist:innen und anderen Beteiligten aus der bisherigen Arbeit in PSI-Debattenräumen ergeben, und ordnet die Befunde medienethisch ein.

Janis Brinkmann geht allgemeiner der Frage nach, wie sich journalistische Qualität unter Bedingungen eines „entgrenzten“ und ausdifferenzierten Journalismus neu bestimmen lässt. Ausgangspunkt ist die Diagnose, dass klassische, am Informations- und Nachrichtenjournalismus orientierte Qualitätsbegriffe gegenüber alternativen Berichterstattungsmustern wie etwa dem investigativen, narrativen oder konstruktiven Journalismus an Grenzen stoßen. Auf öffentlichkeitstheoretischer Grundlage plädiert er für ein flexibles Set von Qualitätskriterien mit einem festen Kern (Aktualität, Relevanz,

Faktizität, Unabhängigkeit), das je nach Journalismus-Typ um spezifische Kriterien ergänzt wird. So soll Qualitätsbewertung gegenstandsadäquater, transparenter und inklusiver werden, ohne in Beliebigkeit abzugleiten.

Danksagung

Hervorgegangen ist dieser Band aus der 10. Jahrestagung der Fachgruppe Kommunikations- und Medienethik der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPK) in langjähriger Kooperation mit dem Netzwerk Medienethik, dem Zentrum für Medien und die digitale Gesellschaft und der Akademie für Politische Bildung, die im Februar 2025 erstmals an der Akademie für Politische Bildung in Tutzing stattfand. Wir danken den Sprecher:innen der Fachgruppe Kommunikations- und Medienethik, Lars Rademacher und Claudia Paganini, dem Vorbereitungsteam der Tagung sowie allen Beteiligten an der Tagung. Insbesondere danken wir den Autor:innen des Bandes für ihre Beiträge. Besonderer Dank gebührt auch Kristína Janačková, Patrick Lessmeister, Khurshida Tarik und Philipp Staber bei der Unterstützung der redaktionellen Publikationsarbeit. Dem Nomos-Verlag danken wir für die Aufnahme des Bandes in die Reihe.

Literatur

- Aïmeur, Esma / Amri, Sabine / Brassard, Gilles* (2023): Fake news, disinformation and misinformation in social media: a review, in: *Social Network Analysis and Mining*, 1–36, <https://doi.org/10.1007/s13278-023-01028-5>
- Debatin, Bernhard* (1997): Ethische Grenze oder Grenze der Ethik? Überlegungen zur Steuerungs- und Reflexionsfunktion der Medienethik, in: Günter Bentele / Michael Haller (Hg.): *Aktuelle Entstehung von Öffentlichkeit. Akteure – Strukturen – Veränderungen*. Konstanz, S. 281–290.
- Muñoz, Katja* (2024): Gegen den Strich: Künstliche Intelligenz und Wahlen, in: *Internationale Politik*, Ausg. 4, S. 36–41.
- Riedel, Manfred* (1979): *Norm und Werturteil*, Stuttgart.
- Ruppert-Karakas, Sascha* (2024): Die Politik des Zorns. Wie die Vordenker der Neuen Rechten den Umsturz vorbereiten, in: *Blätter* 5/2024, S. 89–98, (online unter: <https://www.blaetter.de/ausgabe/2024/mai/die-politik-des-zorns> – letzter Zugriff 28.01.2026).

I.

Umstrittene Sprache: Hate Speech und Gender-Diskurse

Was ist Hassrede? Versuch einer Schärfung eines umstrittenen Begriffs

Inga Bones

Zusammenfassung

Der Begriff Hassrede ist im öffentlichen Diskurs in den vergangenen Jahren allgegenwärtig geworden. Empirische Studien zeigen deutlich die Dimension ihrer Verbreitung und im politischen und regulatorischen Bereich werden Gegenmaßnahmen gefordert. Es wird jedoch seltener gefragt, was „Hassrede“ genau ist. Das führt zu einer uneinheitlichen Verwendung, die auch die Einordnung entsprechender Studienergebnisse und die Evaluation von Maßnahmen erschwert. Der Beitrag diskutiert kritisch einschlägige Begriffsbestimmungen und zeigt, dass Hassrede als Form der Kommunikation zu verstehen ist, die eine Person oder Gruppe auf Basis stabiler und identitätsrelevanter Eigenschaften wie Religion, Ethnie, Geschlecht oder Herkunft auf tiefgreifende Weise angreift, abwertet oder diskriminiert.

1. Einleitung

Hassrede (englisch *hate speech*) und ihre Bekämpfung waren in den vergangenen Jahren immer wieder Thema in den deutschsprachigen Medien. Anlässe, Hassrede in den Fokus zu nehmen, gab es bedauerlicherweise viele: Der Kasseler Regierungspräsident Walter Lübcke etwa wurde vor seiner Ermordung durch einen Rechtsextremisten im Jahr 2019 zur Zielscheibe von hasserfüllten Kommentaren, Gewaltaufrufen und Morddrohungen. In den vergangenen Bundestagswahlkämpfen mussten sich zahlreiche Politiker:innen mit Hasskommentaren auseinandersetzen (vgl. Hoppenstedt 2021), während der Corona-Pandemie wurde gegen Wissenschaftler:innen wie Christian Drosten oder den Gesundheitsminister Karl Lauterbach gehetzt (vgl. Charisius 2021) und seit dem Anschlag der Hamas auf Israel am 7. Oktober 2023 ist ein Anstieg antisemitischer Hassrede zu verzeichnen (vgl. Hoppenstedt 2023, Rech 2024). Und immer wieder ist von Umfrageergebnissen zu lesen, denen zufolge ein großer Teil aller Internetnutzer:in-

nen, insbesondere jüngerer, schon einmal mit Hassrede konfrontiert oder sogar selbst von ihr betroffen war (vgl. Daniel 2023).

Vor dem Hintergrund solcher Meldungen drängt sich die Frage auf, was wir als Gesellschaft gegen Hassrede unternehmen können: Welche Maßnahmen zur Eindämmung von Hassrede sollten staatliche Institutionen, die Betreiber sozialer Netzwerke, die Anbieter von Messengerdiensten (und so weiter) ergreifen? Angesichts der Vielzahl der in den Medien verwendeten Ausdrücke – *Hasskommentar*, *Onlinehetze*, *Beschimpfung*, *Bedrohung*, *Beleidigung*, *fake news* – stellt sich aber auch eine andere, und grundlegendere, Frage: Was genau ist Hassrede eigentlich? Dieser Frage möchte ich in diesem Text nachgehen.

2. Woher stammt und wer gebraucht den Ausdruck „Hassrede“?

2.1 Woher stammt der Ausdruck „Hassrede“?

In seiner Verwendung als Kollektivum ist der Ausdruck „Hassrede“ eine Lehnübersetzung aus dem Englischen („hate speech“) und ein relativer Neuzugang des deutschen Alltagswortschatzes. In der deutschsprachigen (Buch-) Literatur tauchte er vor der Jahrtausendwende kaum auf. Zwischen 2000 und 2010 nahm die Häufigkeit seiner Verwendung zu, um schließlich in den 2010er Jahren sprunghaft anzusteigen. Dieser sprunghafte Anstieg ist für denselben Zeitraum auch für den Ausdruck *hate speech* in der englischsprachigen (Buch-) Literatur zu verzeichnen.¹ Eine Recherche mit dem *Google Books Ngram Viewer* zeigt aber auch, dass der Terminus *hate speech* sich im englischen Sprachraum früher als im deutschen zu verbreiten begonnen hat, nämlich bereits in den späten 1980er und frühen 1990er Jahren. Das deckt sich mit einer Untersuchung von Alexander Brown (2017: 424), demzufolge der Ausdruck *hate speech* von einer Gruppe US-amerikanischer Jurist:innen und Aktivist:innen der Post-Bürgerrechtsära nach 1965 geprägt wurde – und zwar in den achtziger und frühen neunziger Jahren.

Wegweisend war hier ein 1989 publizierter Aufsatz der Rechtswissenschaftlerin Mari Matsuda über rassistisch motivierte Beleidigungen und Drohungen. Matsudas Text basiert auf einem Vortrag, den sie ein Jahr zuvor auf einer Konferenz zum Spannungsverhältnis von Redefreiheit und

1 Die hier besprochenen Daten zur Verwendungshäufigkeit der Termini „Hassrede“ und „hate speech“ basieren auf einer Suche mit dem *Google Books Ngram Viewer*.

sprachlicher Gewalt an der *Hofstra University* gehalten hatte, und enthält ganze vierzig Vorkommnisse des Ausdrucks *hate speech*. 1993 schließlich erschien mit dem Sammelband *Words That Wound* ein weiterer für die gesellschaftliche Debatte um Hassrede zentraler Text, der unter anderem Beiträge von Matsuda, Richard Delgado und Kimberlé Crenshaw versammelt. Alle drei gehören auch zu den Begründer:innen der *critical race theory*, die sich zum Ziel gesetzt hatte, für strukturellen Rassismus und die Intersektionalität von Diskriminierung zu sensibilisieren. Die Verbreitung des Ausdrucks *hate speech* signalisiert also auch ein gesteigertes öffentliches Bewusstsein für gesellschaftliche Machtverhältnisse, strukturelle Diskriminierung und für solche Phänomene wie Alltagsrassismus.

2.2 Ist „Hassrede“ ein umkämpfter Begriff?

Der Ausdruck „Hassrede“ ist das, was man in der Sprachphilosophie einen Hybridausdruck nennt: ein Ausdruck, der nicht nur beschreibt, sondern auch eine Wertung enthält. Wer Hassrede äußert, tut per se etwas Schlechtes – anders als jemand, der eine Feststellung macht oder eine Befürchtung äußert. Wenn wir einen Sprechakt als Hassrede klassifizieren, dann machen wir deutlich, dass wir ihn für verwerflich halten, für etwas, dessen moralische oder rechtliche Sanktionierung im öffentlichen Interesse liegt. Letzteres erklärt auch, warum der Begriff der Hassrede umkämpft ist: Die Freiheit der Meinungsäußerung ist ein schützenswertes Gut und ein Grundrecht, das nicht nur in unserer deutschen Verfassung, sondern auch in der europäischen Charta der Grundrechte verankert ist. Welche Sprechakte eine Gesellschaft sanktionieren sollte, ist deshalb Gegenstand kontroverser Debatten. Eine von Verfechtern einer möglichst weitgehenden Redefreiheit manchmal geäußerte Sorge ist, dass Sanktionen von (vermeintlich oder tatsächlich) problematischen Sprechakten zu einem sogenannten *chilling effect* führen könnten, einer Form der Selbstzensur durch Personen, die zwar eine streitbare, aber durchaus keine extremistische politische Meinung vertreten (vgl. Simpson 2024). Andererseits, so eine mögliche Replik, droht eine Zunahme von Hassrede gerade diejenigen Menschen „mundtot“ zu machen (englisch *to silence*), die sich aufgrund ihrer Zugehörigkeit zu Minderheitengruppen im gesellschaftlichen Diskurs ohnehin nur schwer Gehör verschaffen können (vgl. West 2012).

2.3 Wer gebraucht den Ausdruck „Hassrede“?

„Hassrede“ ist kein juristischer Begriff. Doch obwohl er im Strafgesetzbuch nicht vorkommt, weist er einige Bezüge zu juristischen Tatbeständen auf. Ein zentraler Bezugspunkt ist die Volksverhetzung, andere sind die sogenannten Ehrdelikte der Beleidigung, Verleumdung und üblen Nachrede. Trotz dieser Bezugspunkte ist es nicht sinnvoll, „Hassrede“ als simplen Sammelbegriff für die genannten Straftatbestände zu verstehen – mehr dazu später.

In Ermittlungsbehörden wie den Landespolizeien wurde im Jahr 2001 durch Beschluss der Innenministerkonferenz ein bundesweites Klassifikationssystem zur Erfassung politisch motivierter Kriminalität (PMK) eingeführt. Gleichzeitig wurde mit dem Oberthemenfeld „Hasskriminalität“ eine Möglichkeit geschaffen, solche Straftaten gesondert zu erfassen, die „durch gruppenbezogene Vorurteile motiviert begangen“ werden. Als gruppenbezogene Vorurteile gelten dabei „Vorurteile des Täters bezogen auf Nationalität, ethnische Zugehörigkeit, Hautfarbe, Religionszugehörigkeit, sozialen Status, physische und/oder psychische Behinderung oder Beeinträchtigung, Geschlecht/geschlechtliche Identität, sexuelle Orientierung, [und] äußeres Erscheinungsbild“ (Bundeskriminalamt 2024). Hasskriminalität umfasst jedoch deutlich mehr als bloße Hassrede, so etwa Sachbeschädigungen, aber auch Gewalttaten wie Körperverletzung oder Mord. Am nächsten kommt dem, was wir intuitiv unter Hassrede verstehen, vermutlich das, was in den Statistiken des Bundesinnenministeriums und des BKA als „politisch motivierte Kriminalität mit dem Tatmittel Hassposting“ (ebd.) auftaucht. Allerdings ist diese Kategorie zu eng, weil ihr zufolge Hassrede auf öffentlich zugängliche Beiträge im Internet beschränkt wäre. Hassrede ist aber kein reines Netzphänomen, auch wenn die Besonderheiten digitaler Räume – zum Beispiel die weitgehende Anonymität – Hassrede zu befeuern scheinen.

Anders als Rechts- und Ermittlungsbehörden verwenden Behörden wie die Bundeszentrale für politische Bildung, die Landesmedienanstalten und zivilgesellschaftliche Organisationen die Ausdrücke „Hassrede“ und „hate speech“ zwar explizit, aber uneinheitlich. Dies führt unter anderem zu sehr unterschiedlichen Untersuchungsergebnissen, wenn es um die Frage geht, welcher Anteil der Bevölkerung bereits mit Hassrede konfrontiert war. Eine Forsa-Umfrage im Auftrag der Landesanstalt für Medien NRW im Jahr 2024 ergab zum Beispiel, dass bereits 78 Prozent aller Internetnutzer:innen

ab 14 Jahre schon einmal mit Hassrede im Netz konfrontiert waren, 42 Prozent sogar „häufig“ oder „sehr häufig“ (Landesanstalt für Medien NRW 2024).² Selbst betroffen war laut dieser Studie rund ein Viertel aller Befragten. Eine andere Studie im Auftrag des Vereins Campact und des Instituts für Demokratie und Zivilgesellschaft (IDZ) legt nach eigener Auskunft eine engere Definition von Hassrede zugrunde und kommt zu deutlich kleineren Fallzahlen: Nur 40 Prozent der Befragten haben *hate speech* wahrgenommen, 8 Prozent waren selbst betroffen. Im Unterschied zur ersten Studie waren die Teilnehmer:innen dieser zweiten Studie vorab dazu angehalten worden, Hassrede von individuellen – also nicht gruppenbezogenen – Beleidigungen, Belästigungen oder *cybermobbing* zu unterscheiden (Geschke et al. 2019).

Der Ausdruck „Hassrede“ wird von Behörden, zivilgesellschaftlichen Organisationen und den Medien also höchst uneinheitlich verwendet; er wird oftmals nicht abgegrenzt von *fake news* und *cybermobbing*, individuellen Beleidigungen oder Belästigungen; und manchmal als reines Netzphänomen definiert. Welches Phänomen haben wir – als Gesellschaft, als Medienschaffende, als Angehörige staatlicher und nicht-staatlicher Institutionen – also eigentlich im Blick, wenn wir von Hassrede sprechen? Mit Götz Hamann (2024) müssten wir wohl antworten: Gar keines, oder zumindest kein Konkretes. In einem Artikel aus dem letzten Jahr schreibt Hamann: „Es gibt politische Begriffe, die waren gestern spektakulär und sind heute ziemlich leer. [...] Medien [nutzen] die Worte Hass und Hetze viel zu häufig als bequeme Metapher in Überschriften, was dazu führt, dass Hass und Hetze den Status einer Füllmasse erreicht haben. Sie sind oft nur noch rhetorisches Moltofill, kitten Probleme eher zu, als sie konkret zu benennen. Der Begriff Hassrede ist auf demselben Weg.“ Hamann verbindet diese Diagnose mit der Forderung, auf die Verwendung der genannten Begriffe zu verzichten, sie zu „streichen, wo es nur geht.“ Er fährt fort: „Wir haben an vielen Stellen schon die passenden Begriffe: Volksverhetzung, Verleumdung, Androhung von Gewalt, Rassismus.“

Anstatt den Begriff der Hassrede aus unserem Wortschatz zu streichen, könnten wir als Antwort auf seine uneinheitliche und vielleicht tatsächlich inflationäre Verwendung auch versuchen, ihn zu schärfen – und klar abzugrenzen von den Begriffen, die Hamann aufzählt. Die Frage wäre dann

2 „Wie häufig haben Sie persönlich schon Hate Speech bzw. Hasskommentare im Internet gesehen – zum Beispiel auf Webseiten, in Blogs, in sozialen Netzwerken oder in Internetforen?“

weniger, welches Phänomen wir im Blick haben, wenn wir den Ausdruck „Hassrede“ verwenden, sondern welches Phänomen wir im Blick haben *sollten*. Oder, anders formuliert: Die Frage wäre weniger, was Hassrede *ist*, als was Hassrede *sein sollte*. Bei der Beantwortung dieser (normativen) Frage können wir uns an paradigmatischen Beispielen orientieren, die in den Rechtswissenschaften und der Philosophie diskutiert wurden.

3. Was ist Hassrede? Oder: Wie sollten wir Hassrede verstehen?

3.1 Hassrede und Hass

Fragen wir uns also, was wir denn nun eigentlich unter Hassrede verstehen sollten, dann liegt vielleicht eine Antwort auf der Hand: nämlich, dass Hassrede „der sprachliche Ausdruck von Hass“ (Meibauer 2013) sei. Wer schriftlich oder mündlich Hassrede äußert, so die Idee, der verleiht seinem Hass mit sprachlichen Mitteln Ausdruck. Diese naheliegende Antwort ist aber in mehrerlei Hinsicht problematisch: Erstens scheint Hassrede mehr zu umfassen als nur schriftliche oder mündliche Äußerungen, nämlich auch nicht-sprachliche Symbole und Bildmaterial sowie Gesten. In dem eingangs genannten Aufsatz der Rechtswissenschaftlerin Mari Matsuda etwa wird das Anbringen eines Swastika-Symbols (Hakenkreuz) an den Schreibtischen einiger afro-amerikanischer Mitarbeiter der Feuerwehr in San Francisco als Beispiel rassistischer Hassrede angeführt und die US-amerikanische *Anti Defamation League* führt das OK-Handzeichen seit einiger Zeit als Hasssymbol der Ideologie weißer Vorherrschaft. Zweitens kann man alles Mögliche hassen, zum Beispiel den gegnerischen Fußballverein oder verkochte Spaghetti. Aber „Verdammter HSV!“ oder „Ekelerregernder Nudelmatsch!“ würden wir nicht als Hassrede klassifizieren. Seinem Hass sprachlich oder auf anderem Wege Ausdruck zu verleihen, kann also keine hinreichende Bedingung für das Vorliegen von Hassrede sein. Und drittens kann auch Hassrede äußern, wer keinen Hass verspürt: Vielleicht hetzt jemand aus politischem Kalkül gegen bestimmte gesellschaftliche Gruppen oder aus falsch verstandener Loyalität gegenüber Personen aus der eigenen Peergroup. Folglich sind Hassgefühle auf Seiten des Sprechers oder der Sprecherin auch keine notwendige Bedingung für das Vorliegen von Hassrede.

Teresa Marques (2023) argumentiert in einem Aufsatz für die These, dass Hassrede in einem bestimmten Sinn aber sehr wohl „der Ausdruck von

Hass“ sei – in dem Sinne nämlich, in dem das Vorliegen einer Emotion des Hasses auf Sprecher:innenseite zu den konventionell festgelegten Angemessenheitsbedingungen für die Äußerung von Hassrede gehört. Ein Beispiel zur Illustration des Grundgedankens geht auf David Kaplan (2004) zurück: Eine Äußerung wie „Autsch!“ drückt Kaplan zufolge aus, dass die Person, welche die Äußerung vornimmt, eine Schmerzempfindung hat. Die Verwendung des Ausdrucks „Autsch!“ ist in unserer Sprache so geregelt, dass sie nur dann *expressiv korrekt* ist, wenn die Sprecher:in tatsächlich einen Schmerz verspürt. Dies schließt den Fall nicht aus, dass die Sprecher:in den Ausdruck unaufrichtig verwendet – entscheidend ist, dass „Autsch!“ ein konventionelles sprachliches Werkzeug für den Ausdruck von Schmerz ist. Ganz analog, so Marques, ist Hassrede ein konventionelles sprachliches Werkzeug für den Ausdruck von Hass. Wenn jemand rein aus politischem Kalkül hetzt, dann verwendet er oder sie Hassrede zwar unaufrichtig, aber die geäußerten Worte drücken dennoch per semantische Konvention Hass aus.

Marques' These ist plausibel, sofern der Fokus auf sogenannten *slurs* liegt: Herabwürdigende Gruppenbezeichnungen wie das N-Wort oder andere Ethnophaulismen sind offenbar tatsächlich semantisch markiert und werden auch dann als diskriminierend wahrgenommen, wenn die Sprecher:innen keine negativen Emotionen gegenüber der bezeichneten Gruppe verspüren. Aber Hassrede umfasst deutlich mehr und komplexere Äußerungen. Es ist weniger plausibel, dass grammatisch komplexe und häufig stark kontextgebundene Bildungen wie „Illegale Einwanderer in ihre Herkunftsländer abschieben!“ auf semantischer Ebene als hass-expressive Ausdrücke markiert sind.

Wenn Hass eine zentrale Rolle in einer Definition von Hassrede spielen soll, dann vielleicht eher in dem Sinne, dass Hassrede *zu Hass aufstacheln* soll – Hass wäre dann also nicht die tatsächliche Motivation des Sprechers oder der Sprecherin oder die von der Äußerung ausgedrückte Emotion, sondern der intendierte Effekt des Sprechakts. Eine solche Auffassung rückt den „Hass“ in „Hassrede“ dann in die Nähe der Volksverhetzung. Und tatsächlich sind die Ausdrücke „Hass“ und „Hetze“ etymologisch eng verwandt: Im Germanischen bedeutet das Verb „hassen“ so viel wie „feindlich verfolgen“. Diese Ursprungsbedeutung ist in den Wörtern „Hatz“ – eine Variante der Hetzjagd – und „Hetze“ erhalten geblieben. Für ein Verständnis von Hassrede als Volksverhetzung spricht auch, dass Hassrede von hochrangigen Stellen als Vorläufer von Verbrechen im Sinne des Völkerrechts und insbesondere von Genoziden charakterisiert wird, etwa von Antonio

Guterrez (2019) anlässlich der Vorstellung des Aktionsplans der Vereinten Nationen gegen *hate speech*.

Guterrez erinnert in seiner Rede daran, dass beispielsweise dem Völkermord an den Tutsi in Ruanda eine systematische Hetzkampagne in an die Hutu adressierte Zeitungs- und Radiopropaganda vorausging. Auch Lynne Tirrell (2012), eine US-amerikanische Sprachphilosophin, zieht die Propaganda gegen die Tutsi, die öffentlich als „inyenzi“, Kakerlaken, entmenslicht wurden, als ein paradigmatisches Beispiel für Hassrede heran. Aber: Wenn man Hassrede (ausschließlich) als Volksverhetzung versteht, bleiben andere paradigmatische Fälle außen vor. Volksverhetzung liegt nämlich erst dann vor, wenn eine Äußerung geeignet ist, den öffentlichen Frieden zu stören. Das ist aber erst dann der Fall, wenn diese Äußerung ein gewisses Ausmaß der Verbreitung erreicht und wenn sie „ihrem Inhalt nach erkennbar auf rechtsgutgefährdende Handlungen hin angelegt [ist], d.h. den Übergang zu Aggression oder Rechtsbruch markier[t].“³ Eine ganz ähnliche Auffassung vertritt bereits John Stuart Mill in *On Liberty*. Mill (1974/1859: 77) schreibt: „Selbst Gedanken verlieren ihre Strafflosigkeit, wenn die Umstände, unter denen sie ausgesprochen werden, von der Art sind, daß ihr Ausdruck eine direkte Aufreizung zu irgendeiner Schandtat bildet.“ Ausgeschlossen wären damit Fälle, in denen Einzelpersonen direkt – und nicht im Beisein eines größeren Publikums – verbal angegriffen, bedroht oder eingeschüchtert würden, aber auch Fälle, die nicht erkennbar auf rechtsgutgefährdende Handlungen angelegt sind. In einer anderen Hinsicht scheint die Kategorie der Volksverhetzung aber auch zu weit zu sein. Noch einmal Mill (ebd.): „Die Meinung, daß Getreidehändler die Armen aushungern [...] sollte gerechterweise Strafe nach sich ziehen, wenn man sie mündlich einer erregten Menge, die sich vor dem Hause eines Getreidehändlers versammelt hat, vorträgt oder sie unter gleichen Umständen in Form von Handzetteln in Umlauf setzt.“ Es sind Umstände denkbar, unter denen Mills Beispiel „Getreidehändler hungern die Armen aus“ zu Gewalttaten aufstacheln und damit den Straftatbestand der Volksverhetzung erfüllen würde. Ob es sich aber um Hassrede handelt, ist zumindest fraglich.

Es scheint unstrittig, dass jede beliebige gesellschaftliche Gruppe unter bestimmten Umständen den Hass – oder, vorsichtiger, die Abneigung – anderer Teile der Bevölkerung auf sich ziehen kann. Prinzipiell kann gegen Getreidehändler:innen, gegen Schulbusfahrer:innen und Grundschul-

3 Aus einer Grundsatzentscheidung des Bundesverfassungsgerichts im Jahr 2009, der sogenannten Wunsiedel Entscheidung.

lehrer:innen, gegen SUV-Fahrer:innen und die Besitzer:innen von Gucci-Handtaschen gehetzt werden. Dass uns solche Gruppen eher selten in den Sinn kommen, wenn wir von Hassrede hören oder lesen, hat sicherlich etwas damit zu tun, dass eine grundsätzliche Abneigung gegen diese Gruppen nicht sehr verbreitet ist. Vielleicht spielt aber auch eine Rolle, dass die Eigenschaften, Schulbusse zu fahren oder Gucci-Handtaschen zu besitzen, nicht so viel damit zu tun haben, *wer wir sind* – dass sie nicht stabil, nicht zentral für unsere Identität, sondern verlierbar und zufällig sind.

3.2 Hassrede und gruppenbezogene Vorurteile

Erinnern wir uns noch einmal an die Anfänge des Ausdrucks „hate speech“. Der Ausdruck wurde vor dem Hintergrund von größer angelegten Bestrebungen geprägt, strukturelle Diskriminierung sichtbar zu machen und zu kritisieren. Im Fokus standen dabei solche gruppenspezifischen Eigenschaften, die stabil und zentral für unsere Identität sind und an denen sich Diskriminierung historisch gesehen besonders häufig festgemacht hat: Allen voran „race“, aber auch die ethnische Herkunft, Nationalität, Klasse, Geschlecht oder Gender, Religion oder sexuelle Orientierung. Manchmal werden diese Eigenschaften auch „protected characteristics“ genannt – und in Artikel 3 unseres Grundgesetzes werden die meisten von ihnen explizit genannt.

Der UN-Aktionsplan gegen Hassrede nimmt diese besonderen Schutz verdienenden Eigenschaften in seine Definition mit auf: Hassrede ist „jede Form der Kommunikation in Wort, Schrift oder Verhalten, die eine Person oder eine Gruppe auf der Basis dessen, was sie sind – mit anderen Worten, auf der Basis ihrer Religion, Ethnie, Nationalität, „Rasse“, Hautfarbe, Abstammung, Gender – angreift, abwertet oder diskriminiert.“⁴ Diese Definition umfasst sowohl paradigmatische Fälle von *Hassrede als Volksverhetzung* – wie etwa die Hetzkampagne gegen die Tutsi – als auch paradigmatische Fälle von *Hassrede als Beleidigung* oder *Bedrohung*, die sich direkt an das Opfer wendet – etwa, wenn einer *Person of Color* auf offener Straße das N-Wort entgegengeschleudert wird. Als Richtschnur dafür, welche Sprechakte wir rechtlich oder sozial sanktionieren sollten, scheint die Definition

4 “[...] any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.”

aber (erneut) zu weit: Äußerungen wie „Mädchen können nicht Fußball spielen“ oder „Deutsche haben keinen Geschmack“ sind diskriminierend und abwertend. Als Hassrede würden sie vermutlich die wenigsten von uns bezeichnen wollen. Damit ein kommunikativer Akt die Bezeichnung als „Hassrede“ verdient, muss offenbar ein gewisser Schweregrad gegeben sein. Sonst verliert der Begriff, wie von Hamann befürchtet, wohl tatsächlich an „politischer Wucht“. Wann ein hinreichender Schweregrad erreicht ist, lässt sich nicht präzise ausbuchstabieren und ist sicher auch Gegenstand gesellschaftlicher Aushandlung. Als Orientierung können uns aber eben jene paradigmatischen Beispiele dienen, die in der wissenschaftlichen Literatur zu *hate speech* wiederholt auftauchen und von denen oben einige angeführt wurden.

Wenn Hassrede die Zugehörigkeit zu einer durch geschützte Eigenschaften definierten Gruppe voraussetzt, dann kann es keine Hassrede gegen Politiker:innen *als Politiker:innen* oder gegen Wissenschaftler:innen *als Wissenschaftler:innen* geben. „Alle Politiker:innen gehören an den Pranger“ oder „Weg mit dem korrupten Wissenschaftlerpack“ qualifizieren sich also nicht als Hassrede. (Aber sie können sich unter bestimmten Umständen durchaus als Volksverhetzung qualifizieren.) Trotzdem kann es sein, dass Politiker:innen und andere Personen des öffentlichen Lebens besonders häufig zu Opfern von Hassrede werden, nämlich dann, wenn sie besonders häufig auf der Basis ihres Geschlechts, ihrer Herkunft, ihrer religiösen Überzeugungen oder ihrer sexuellen Orientierung angefeindet werden.

4. Fazit

„Hassrede“ ist ein umkämpfter Begriff, der in der öffentlichen Debatte, von Behörden und Organisationen uneinheitlich verwendet wird. Ich habe für eine Schärfung des Begriffs argumentiert, die sich weniger an der expressiven Funktion von Hassrede als an ihrer Verknüpfung mit gruppenbezogenen Vorurteilen orientiert. Demnach ist Hassrede eine Form der Kommunikation, die eine Person oder Gruppe auf der Basis stabiler und identitätsrelevanter Eigenschaften (wie Religion, Ethnie, Geschlecht oder Herkunft) in schwerwiegender Weise angreift, abwertet oder diskriminiert.

Eine Schärfung des Begriffs der Hassrede und damit eine klare Abgrenzung von persönlichen Beleidigungen ist insbesondere dann relevant, wenn empirische Forschung belastbare Daten zur Prävalenz und zu den Auswirkungen von Hassrede – zum Beispiel in den sozialen Medien – liefern

soll. Auch die (manuelle oder automatisierte) Moderation von Beiträgen in sozialen Medien profitiert von einem geschärften Begriffsverständnis, etwa wenn im Einzelfall über die mögliche Überschreitung der Grenze zur Strafbarkeit oder über die juristische Grundlage einer Strafverfolgung entschieden werden muss.

Literatur

- Bundeskriminalamt* (2024): Definitionssystem. Politisch motivierte Kriminalität (online unter: https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/definitionssystem-pmk.pdf?__blob=publicationFile&v=2 – letzter Zugriff: 12.1.2026).
- Brown, Alexander* (2017): What is hate speech? Part 1: The myth of hate, in: *Law and Philosophy* 36, S. 419–468.
- Charisius, Hanno* (2021): Jeder siebte Corona-Experte berichtet von Morddrohungen, in: *Süddeutsche Zeitung*, 13. Oktober 2021 (online unter: <https://www.sueddeutsche.de/gesundheitscorona-forscher-drohungen-1.5438622> – letzter Zugriff: 21.9.2025).
- Daniel, Isabelle* (2023): Mehr als jeder vierte Internetnutzer hat Erfahrung mit Hassrede, in: *Die ZEIT*, 11. Dezember 2023 (online unter: <https://www.zeit.de/gesellschaft/zeitgeschehen/2023-12/statistisches-bundesamt-hatespeech-internet> – letzter Zugriff: 21.9.2025).
- Geschke, Daniel et al.* (2019): #Hass im Netz: Der schleichende Angriff auf unsere Demokratie. (online unter: https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf – letzter Zugriff: 12.1.2026).
- Gutierrez, Antonio* (2019): Secretary-General's Remarks at the Launch of the United Nations Strategy and Plan of Action on Hate Speech (online unter: <https://www.un.org/sg/en/content/sg/statement/2019-06-18/secretary-generals-remarks-the-launch-of-the-united-nations-strategy-and-plan-of-action-hate-speech-delivered> – letzter Zugriff: 21.9.2025).
- Hamann, Götz* (2024): "Hass und Hetze" klingt heute hohl, in: *Die ZEIT*, 20. Mai 2024 (online unter: <https://www.zeit.de/kultur/2024-05/politische-sprache-hass-hetze-bedrohung-kritik> – letzter Zugriff: 21.9.2025).
- Hoppenstedt, Max* (2021): Täglich Hass, tausendfach, in: *Spiegel Online*, 20. September 2021 (online unter: https://www.spiegel.de/netzwelt/netzpolitik/hasskommentare-gegen-die-kanzlerkandidaten-schimpf-und-schande-a-ff7b6878-f9d7-46e2-b4b9-e0a78a2c4957?sara_ref=re-xx-cp-sh – letzter Zugriff: 21.9.2025).
- Hoppenstedt, Max* (2023): Ermittler verzeichnen deutliche Zunahme antisemitischer Hasskommentare, in: *Spiegel Online*, 6. November 2023 (online unter: https://www.spiegel.de/netzwelt/antisemitismus-ermittler-sehen-deutliche-zunahme-von-online-hasskommentaren-a-ff2f4fa3-4b2c-449c-aca9-49d06fab1ea2?sara_ref=re-xx-cp-sh – letzter Zugriff: 21.9.2025).

- Kaplan, David* (2004): The Meaning of Ouch and Oops, Howison Lecture in Philosophy, UC Berkeley, 23. August 2004 (Transkription online unter: <https://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf> – letzter Zugriff: 21.9.2025).
- Landesanstalt für Medien NRW* (Hg.) (2024): Hate Speech. Forsa-Studie 2024 (online unter: https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/sicherheit/definitionssystem-pmk.pdf?__blob=publicationFile&v=2 – letzter Zugriff 12.1.2026).
- Marques, Teresa* (2023): The Expression of Hate in Hate Speech, in: *Journal of Applied Philosophy* 40, S. 769–787.
- Matsuda, Mari* (1989): Public Response to Racist Speech: Considering the Victim's Story, in: *Michigan Law Review* 87 (8/1989), S. 2320–2381.
- Meibauer, Jörg* (2013): Hassrede – von der Sprache zur Politik, in: Jörg Meibauer (Hg.), *Hassrede/Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion*, Gießener Elektronische Bibliothek, S. 1–16.
- Mill, John Stuart* (1974/1859), *Über die Freiheit*, Stuttgart.
- Rech, David* (2024): Wie sich antisemitische Kommentare nach dem 7. Oktober verändert haben, *Die ZEIT*, 18. April 2024 (online unter: <https://www.zeit.de/gesellschaft/2024-04/antisemitismus-kommentare-studie-israel-decoding-antisemitism> – letzter Zugriff: 21.9.2025).
- Simpson, Robert Mark* (2024): Self-Censorship: The Chilling Effect and the Heating Effect, in: *Political Philosophy* 1 (2/2024), S. 345–380.
- Tirell, Lynne* (2012): Genocidal Language Games, in: Ishani Maitra / Mary Kate McGowan (Hg.), *Speech and Harm: Controversies over Free Speech*, Oxford University Press, S. 174–221.
- West, Caroline* (2012): Words That Silence? Freedom of Expression and Racist Hate Speech, in: Ishani Maitra / Mary Kate McGowan (Hg.), *Speech and Harm: Controversies Over Free Speech*, Oxford University Press, S. 222–248.

„Squatting for Hitler“: Hate Speech, Datenmüll, und generative KI in Zeiten rechtspopulistischer Autokratie

Bernhard Debatin

Zusammenfassung

Mit Beispielen aus dem US-Wahlkampf und von KI-generierter Hassmusik wird analysiert, wie Hassrede und rassistische, antisemitische und misogynie Propaganda durch KI-Plattformen verstärkt wird, und wie dadurch Normalisierungseffekte und *confirmation bias* entstehen. Aus analytischer Sicht geht es hier zunächst um die Frage, inwieweit die Idee der Technik als bloßes Werkzeug einer wirkungsvollen Entwicklerverantwortung im Wege steht. KI ist ja, im Unterschied zu anderen Technologien, in der Lage, selbst Zwecke zu setzen, also stets mehr als nur Werkzeug. Dies erzeugt das bekannte Problem der *tragedy of the commons* und folglich die zunehmende Vermüllung des Internets mit KI-Content. Zum Abschluss wird diskutiert, wie in den USA unter Präsident Trump unregulierte und ungebremste KI als Propagandawerkzeug instrumentalisiert und normalisiert wird.

1. Die neue politische Landschaft

In den USA findet seit dem Amtsantritt von Präsident Trump der Übergang zur offen oligarchischen Autokratie statt, mit einer Flut von Präsidentenverfügungen (*executive orders*), die an der Gesetzgebung vorbei tiefgreifende Veränderungen bewirken. Viele von ihnen sind fragwürdig, wenn nicht gar illegal, da sie ohne Legitimation Macht- und Entscheidungsbefugnisse der Legislative, zum Beispiel in Haushalts-, Wirtschafts- und Verteidigungsfragen, in die Exekutive verlagern. Dies ist Teil eines nie zuvor gekannten *powergrab*, einer Machtergreifung durch einen Präsidenten mit diktatorischen Ambitionen.

Die MAGA-Ideologien, allen voran Stephen Miller, Russell Vought und ihre Mitstreiter von der ultrakonservativen *Heritage Foundation*, sind – im Unterschied zur ersten Präsidentschaft von Trump – sehr gut und de-