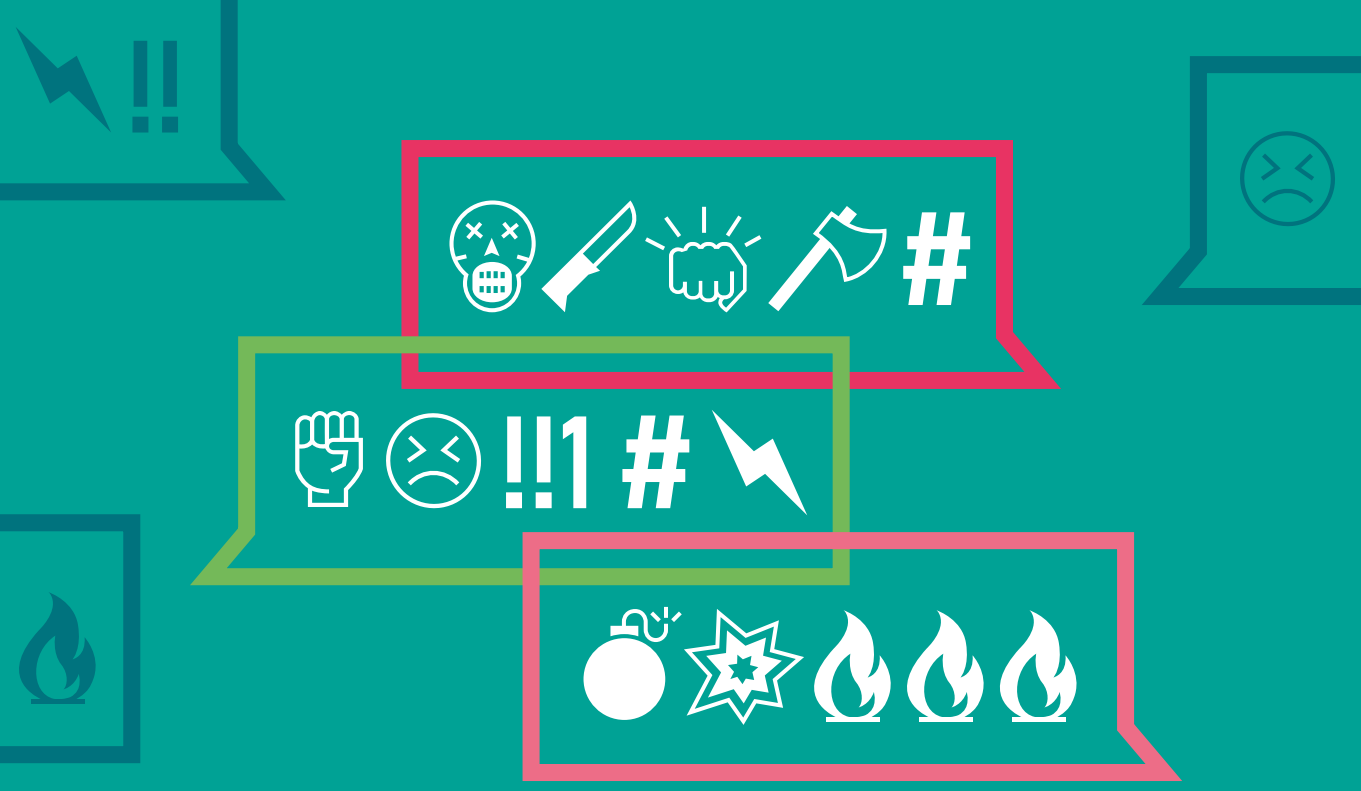


HASSKOMMENTARE IM NETZ.
STEUERUNGSSTRATEGIEN FÜR REDAKTIONEN

Leif Kramp & Stephan Weichert

LANDESANSTALT FÜR MEDIEN NRW
Der Meinungsfreiheit verpflichtet.

Die ausführlichen Ergebnisse werden im Herbst 2018 im VISTAS Verlag veröffentlicht:

Kramp, Leif und Weichert, Stephan:
Hasskommentare im Netz. Steuerungsstrategien für Redaktionen. Leipzig: 2018
(Schriftenreihe Medienforschung der Landesanstalt für Medien NRW, Bd. 79),
ISBN 978-3-89158-647-1, 26,- €

HASSKOMMENTARE IM NETZ.
STEUERUNGSSTRATEGIEN FÜR REDAKTIONEN

Leif Kramp & Stephan Weichert

Unter Mitarbeit von
Viviane Harkort und Lara Malberger

LANDESANSTALT FÜR MEDIEN NRW
Der Meinungsfreiheit verpflichtet.

INHALT

5 Vorwort

6 STRATEGIEN IM MODERATIONS-PROZESS – HANDREICHUNGEN FÜR REDAKTIONEN

10 UNTERSUCHUNGSDESIGN

11 Diskurstypologie

13 Online-Diskursanalyse: Kernergebnisse (Auszug)

15 REDAKTIONELLE MODERATIONS-STRATEGIEN GEGEN HASSREDE

19 Regulierende Strategie: Dis-Empowerment

22 Bestärkende Strategie: Empowerment

26 ANHANG

26 Leitfäden und Automatisierungstools für Redaktionen im Umgang mit Hassrede

30 Die Autoren

VORWORT

Menschen teilen ihre persönlichen Gefühle und Gedanken, sie trauern und freuen sich gemeinsam, neigen zu spontanen Reaktionen, geben aber auch – intentional oder wider besseren Wissens – ungeprüfte oder gar falsche Informationen weiter. Die digitalen Publikumsbeziehungen, konkret das Kommentarverhalten von Mediennutzerinnen und Mediennutzern, verändern die öffentliche Debattenkultur grundlegend: Potenziell gibt es, besonders bei jungen Zielgruppen, vielfältig verbesserte Teilhabemöglichkeiten an gesellschaftlichen Diskursen sowie an politischer Willens- und Meinungsbildung. Zugleich birgt diese Debattenkultur aber auch hohe Risiken für die demokratische Diskursrationalität.

Für Nachrichtenanbieter und ihre Redaktionen sowie für den Journalismus insgesamt stellt die Dialogisierung mit dem Publikum, insbesondere die Sichtung, Moderation, Prüfung und Freischaltung von Kommentaren, nicht nur eine ideale Gestaltungsmöglichkeit dar, sondern sie ist vor allem eines: frustrierend. Während die Recherche und Publikation geschützter Information zu den Kernaufgaben seriöser Berichterstattung gehören, wird die Moderation von Nutzerdiskursen von Redaktionen in den Kommentarspalten ihrer Nachrichtenangebote – angesichts mangelnder Ressourcen – nach wie vor als unangenehme Zusatzaufgabe empfunden.

Im Auftrag der Landesanstalt für Medien NRW und mit finanzieller Unterstützung von Google Deutschland haben wir das Diskussionsverhalten der Nutzerinnen und Nutzer führender Nachrichtenmarken im Netz und deren konkrete Moderationsstrategien untersucht. Analysiert wurden die redaktionellen Websites bzw. jeweils ein Social-Media-Auftritt von Deutschlandfunk Kultur, RP Online, RTL und Tagesschau.de. Im Zentrum des Erkenntnisinteresses stand die Frage, wie journalistische Medien in Interaktion mit ihrem Publikum durch gezielte Strategien und redaktionelle Steuerungsmechanismen (u. a. Moderation, Community Management, Audience Engagement, Löschpraktiken) Nutzerdiskurse konstruktiv begleiten und ausufernde Debatten regulieren können. Die Studie soll Redaktionen Ansatzpunkte in (potenziell) hassgetriebenen Diskussionen aufzeigen, um diese problemlos in den Redaktionsalltag integrieren zu können.

Wir danken: Viviane Harkort, Lara Malberger, Daniel Moßbrucker, Lisa-Marie Eckardt (Projektmitarbeit); Dr. Nicola Balkenhol und Torben Waleczek (Deutschlandfunk Kultur); Daniel Fiene, Julia Nix, Mathias Schumacher und Henning Bulka (RP Online); Christian Beissel (RTL Interactive); Christina Elmer, Torsten Beeck, Eva Horn, Philip Löwe und Werner Theurich (Spiegel Online); Dr. Kai Gniffke, Christiane Krogmann, Rike Woelk und Sven Königsmann (Tagesschau.de); Sabine Frank und Anika Lampe (Google Deutschland); Patricia Georgiou und Lucas Dixon (Perspective API); Dr. Meike Isenberg, Marie-Franca Hesse und Mechthild Appelhoff (Landesanstalt für Medien NRW).

Die ausführlichen Ergebnisse werden im Herbst 2018 im VISTAS Verlag veröffentlicht.

Bremen/Hamburg im Juni 2018,
Leif Kramp & Stephan Weichert

STRATEGIEN IM MODERATIONSPROZESS – HANDREICHUNGEN FÜR REDAKTIONEN

Die wachsenden Anforderungen an Audience Engagement, Community Development und Plattform Management erfordern ebenso viel Energie, Fingerspitzengefühl wie auch ein stabiles Nervenkostüm. Meist fehlt es Redaktionen schlicht an Ressourcen und Zeit, teils auch an speziellen Techniken und Tools, wenn sie Nutzerdiskurse mit Argumenten versachlichen wollen, statt eine emotionsgeladene Debatte eskalieren zu lassen. Um Hassrede, Hetze, Extremismus, Verunglimpfungen und Ausgrenzungen möglichst wenig Raum zu geben, helfen nicht nur Verhaltensregeln, die auf einem ständigen inter- wie intraredaktionellen Erfahrungsaustausch beruhen. Nutzerkommentare konstruktiv zu fördern und weitgehend konfliktfrei zu moderieren, ist vor allem eine Frage der übergreifenden Strategie im Moderationsprozess: Sollten Social-Media-Redakteurinnen und -Redakteure früh in Diskussionen eingreifen oder sie laufen lassen? Wollen Moderatorinnen und Moderatoren Hassrede löschen oder mit Haterinnen und Hatern diskutieren? Agieren sie als Personen oder als Medienmarke? Wie schützen sich Community-Managerinnen und -Manager vor verbaler Gewalt? Sollten Trolle gesperrt, ihre Kommentare gelöscht oder die betreffenden Personen sogar strafrechtlich verfolgt werden? Wir stellen auf der Grundlage unserer empirischen Erkenntnisse einen **10-Punkte-Plan** gegen Hassrede vor, um die teilweise dysfunktionale Debattenkultur in den Kommentarspalten von Nachrichtenangeboten zu entstören.

10 PUNKTE-PLAN

1 Entschieden moderieren

Um seriös Diskutierende in ihrer Mitwirkung und Argumentation zu bestärken, moderieren Sie klar und bestimmt. Sie müssen nicht stillschweigend hinnehmen, dass Ihnen Hater, Störer und Trolle mit ihren primitiven Parolen das Leben schwermachen. Entschärfen und vertreiben Sie den Sprachterror aus Ihrem Kommentarbereich durch eine beherzte, aber sachliche Moderation, die verdeutlicht, wer bei Ihnen das Hausrecht hat (auch auf Ihren Social-Media-Ablegern).

2 Direkte Ansprache

Melden Sie sich häufig zu Wort, statt sich nur darauf zu konzentrieren, problematische Kommentare zu löschen oder auszublenden. Versuchen Sie, böswillige und beleidigende Nutzerinnen und Nutzer direkt anzusprechen und zur Ordnung zu rufen. Oft reicht es schon, wenn Haterinnen und Hater merken, dass sie beachtet und beobachtet werden, um Diskurse zu zivilisieren. Machen Sie in Ihrer Moderation klar, dass Sie auf Ihrer Website und Ihren Social-Media-Angeboten nur Kommentierende akzeptieren, die sich an die Netiquette der Redaktion halten und einen höflichen, fairen Umgang pflegen.

3 Gegenrede stärken

Besonders Gegenredende, die sich in Ihren Kommentarbereichen klar gegen Hassrede einsetzen, gilt es zu belohnen. Finden Sie die loyalen und engagierten Kommentierenden unter Ihren Nutzerinnen und Nutzern und machen Sie sich mit ihnen gemein bzw. solidarisieren Sie sich mit ihnen. Vertrauen Sie ihnen die Ko-Moderation von Gesprächen an, die zu eskalieren drohen, durch kommunikative Selbstregulierung jedoch frühzeitig geschlichtet werden können. Achten Sie vor allem darauf, wo Aktivistengruppen wie #ichbinhier unterwegs sind, die sich in Diskurse konstruktiv einmischen und engagiert gegen Hass und Hetze vorgehen. Durch gezieltes Empowerment von Gegenrede und konstruktiver Kommunikation können die Selbstheilungskräfte im Netz gesteigert und das Immunsystem der Debattenkultur gestärkt werden.

4 Aktionen gegen Hassrede

Erfinden Sie journalistische Programme, Formate und Veranstaltungen, die den Hass bei den Wurzeln packen. Oftmals sind Unwissen, Unverständnis oder Enttäuschung die Ursachen für emotionsgeladene Pöbeleien. Aktionen wie „Sag's mir ins Gesicht" der Tagesschau sind wertvolle Experimente gegen Hate Speech, die zeigen, dass sich der offene Dialog mit frustrierten Kritikerinnen und Kritikern durchaus lohnen kann. Nur flexible Redaktionsarbeit auf Augenhöhe kann Vorwürfe wie „Staatsmedien" und „Lügenpresse" wirksam entkräften.

5 Hässliches Dominanzgefälle

Machen Sie sich klar, dass eine laute Minderheit die digitalen Diskursräume mit ihren Hassbeiträgen beherrscht, während die Mehrheit der Nutzerinnen und Nutzer schweigt. Durch die Bewusstwerdung dieser Schieflage werden Hass und Hetze im Netz zwar nicht abgebaut, aber schon diese Erkenntnis kann Ihnen helfen, die Verschlimmerung der Debattenkultur als wesentlich harmloser wahrzunehmen. Durch das (zeitnahe) Ausblenden von Beiträgen können Störerinnen und Störer sowie Haterinnen und Hater gezielt isoliert werden.

6 Konstruktiver Journalismus

Wohlfühl- und Flauschjournalismus können die Verrohung der Kommentarkultur im harten News-Geschäft nicht stoppen, wohl aber die regelmäßige Veröffentlichung lösungsorientierter Kommentarbeiträge – vor allem solche mit Bezug zum realen Lebensumfeld der Nutzerinnen und Nutzer. Wissenschaftliche Studien zeigen, dass die engagierte Teilhabe der Userinnen und User umso höher ist, je konstruktiver die Berichterstattung ist. Dies trägt auch insgesamt zur Harmonisierung der Debattenkultur bei, denn Kommentierende regen sich dann weniger über Probleme und krisenhafte Entwicklungen auf, sondern berichten dafür mehr über positive Erfahrungen oder Perspektiven.

7 Mensch-Maschine-Filter

Algorithmen und künstliche Intelligenz können die Beurteilung von Kommentaren durch Menschen nicht ersetzen, sie können diese aber durch Vorfilterung erleichtern. Die automatisierte Kanalisierung von Nutzerfeedback mittels technologischer Systeme, die auf Sprach- oder Syntaxerkennung beruhen, können die schlechten Kommentare aussortieren und gegebenenfalls sogar löschen. Wer täglich mehrere tausend Kommentare pro Tag lesen, moderieren und analysieren muss, wird dies als willkommene Entlastung empfinden, um sich stärker den positiven Beiträgen der Nutzerinnen und Nutzer widmen zu können.

8 Ironie- und zynismusfreie Zone

Eine der größten Herausforderungen in der Kommentarmoderation ist, dass Ironie nicht nur von Computern, sondern auch von vielen Userinnen und Usern nicht als solche erkannt wird. Zu einem gesunden Nutzerdiskurs gehört deshalb, ironisierende Moderationselemente wohlüberlegt einzusetzen und grundsätzlich keine Nutzerinnen und Nutzer, auch nicht die unfreundlichen, zu verspotten – selbst wenn es mitunter viel Selbstdisziplin braucht, sich bei üblen Kommentierenden nicht im Tonfall zu vergreifen. Der innere Schweinehund des Zynikers ist ebenfalls kein guter Ratgeber, weil er oft dazu verleitet, Diskussionen einfach abzuwürgen, statt zu sachlichen Beiträgen zu ermutigen.

9 Ressourcen bereitstellen

Am Dialog mit den Nutzerinnen und Nutzern festzuhalten und ihn weiter zu fördern, um das volle Kreativpotenzial der Kommentarmöglichkeiten für die eigene Medienmarke auszuschöpfen, bedeutet vor allem, die nötige Kapazität und Infrastruktur bereitzustellen. Durch Doppelbesetzungen von Social-Media-Redakteurinnen und -Redakteuren pro Plattform, aber auch durch die gezielte Verzahnung von Kommentarmoderation sowie Autorinnen und Autoren kann es gelingen, die fachlich-inhaltlichen Akzente von Debatten in den Vordergrund zu stellen. Statt sich emotionalen Provokationen hinzugeben, sollte Moderation ressourcenstark auf die Beweggründe von Kommentierenden reagieren.

10 Respekt verschaffen

Gerade in einem rauen Kommunikationsklima kommt es nicht nur darauf an, auf Augenhöhe miteinander zu kommunizieren. Ein Unrechtsbewusstsein auf Nutzerseite zu vermitteln erfordert auch, Wiederholungstäterinnen und Wiederholungstätern ihre Grenzen konsequent aufzuzeigen. Dauerhaft aktive Trolle sowie Haterinnen und Hater, die mit ihren Hasseinträgen ganze Online-Nachrichtenangebote verunreinigen, sollten aus den Kommentarbereichen verbannt und gegebenenfalls strafrechtlich verfolgt werden. Initiativen wie „Verfolgen statt nur Löschen“ der Landesanstalt für Medien NRW setzen wichtige Akzente für Kooperationen zur effektiven Bekämpfung von Hasskriminalität, um sichere und faire Diskurse im Netz zu ermöglichen.

UNTERSUCHUNGSDESIGN

Mit den Online-Redaktionen von Deutschlandfunk Kultur, Rheinische Post Online, RTL und Tagesschau konnten für die Studie vier idealtypische Kooperationspartner von unterschiedlichen Qualitätsmediengattungen mit Redaktionssitz in Nordrhein-Westfalen und Berlin (DLF: Köln/Berlin, RP Online: Düsseldorf, RTL Aktuell: Köln) bzw. in Hamburg (Tagesschau.de) gewonnen werden. Außerdem stand mit Spiegel Online (Hamburg) ein weiterer kompetenter journalistischer Partner im Rahmen der Expertengespräche zur Verfügung, der regelmäßig ein extrem hohes Kommentaraufkommen und eine langjährige Erfahrung in der Bearbeitung von Kommentaren beispielsweise in einschlägigen Foren vorzuweisen hat. Mit ausgesuchten Social-Media-Redakteurinnen und -Redakteuren sowie Verantwortlichen dieser Redaktionen wurden im ersten Untersuchungsschritt einschließlich eines Pretests insgesamt zwölf Expertengespräche geführt sowie im zweiten Schritt ausgewählte Online-Diskursverläufe aus der zweiten Jahreshälfte 2017 untersucht. Im dritten Untersuchungsschritt haben wir im Frühjahr 2018 ein Experiment mit einer der Online-Redaktionen im Sample durchgeführt.

Das Sample der Kommentaranalyse ist das Herzstück der Untersuchung. Es besteht aus insgesamt 24 unterschiedlichen Online-Diskursverläufen zu 16 einschlägigen journalistischen Beiträgen (vgl. Tab. 1, S. 12). Dabei wurden die relevanten Plattform- bzw. Moderationsstrategien der Kooperationspartner aufgegriffen: Bereits in Vorgesprächen mit den Redaktionsverantwortlichen stellte sich heraus, dass beispielsweise in Bezug auf YouTube, Twitter und Instagram seitens der Redaktionen keine nennenswerten personellen Ressourcen auf die Social-Media-Moderation verwendet werden, sondern sich die strategischen Überlegungen vor allem auf die Facebook-Pages konzentrieren. Andere Plattformen wie Twitter, YouTube oder Instagram spielen bei der derzeitigen Entwicklung und/oder Ausübung von Moderationsstrategien, aber auch hinsichtlich der redaktionellen Begleitung von Nutzerdiskursen keine bzw. lediglich eine untergeordnete Rolle. Analysiert wurden deshalb die Nutzerdiskurse bzw. das Kommentaraufkommen auf den Facebook-Ablegern der genannten Medienangebote sowie bei RP Online und Tagesschau.de zusätzlich in den Kommentarbereichen der redaktionseigenen Websites.

DISKURSTYPOLOGIE

Die zu den jeweiligen journalistischen Beiträgen auf den Fan-Pages des betreffenden Mediums geführten Nutzerdiskurse lassen sich in fünf verschiedene Typen unterteilen (vgl. Abb. 1):

Abbildung 1: Typologie der analysierten Nutzerdiskurse

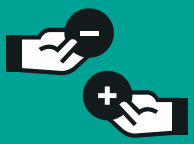
1	Hass- bzw. konfliktgeladene Diskurse zu Beiträgen über gesellschaftspolitische Reizthemen (stark geprägt von vielen destruktiven Kommentaren)	
2	Von starker Negativität und Einzelbeiträgen geprägte Diskurse ohne Eskalationstendenz	
3	Diskurse mit nutzerseitigem Selbstregulierungseffekt (destruktive Kommentare werden durch konstruktive, lösungsorientierte Kommentare neutralisiert)	
4	Diskurse mit überwiegend konstruktiven, lösungsorientierten und/oder bestätigenden Kommentaren (mit viel/wenig Moderationsaktivität der Redaktion)	
5	Diskurse mit lebensnahem Alltagsbezug und hohem, weitgehend neutralem Kommentaraufkommen	

Tabelle 1: Charakterisierung der analysierten Nutzerdiskurse

Diskurscharakterisierung	Medium	Beitragstitel	Veröffentlichung
Hass- bzw. konfliktgeladene Diskurse zu Beiträgen über gesellschaftspolitische Reizthemen (stark geprägt von vielen destruktiven Kommentaren)	RP Online	„Studie bescheinigt Muslimen Erfolge auf dem Arbeitsmarkt“	24.8.2017
	RTL Aktuell	„Flüchtlingsunterkünfte oft mangelhaft“	7.12.2017
	RTL Aktuell	„GEZ-Schock: Neuer ARD-Chef fordert höheren Rundfunkbeitrag“	30.12.2017
	Tagesschau.de	„Milliarden-Deal mit Israel“	23.10.2017
	Tagesschau.de	„Kai Gniffkes Kommentar zur AfD“	12.11.2017
Von starker Negativität und Einzelbeiträgen geprägte Diskurse ohne Eskalationstendenz	RTL Aktuell	„Sexuelle Belästigung: 16 Frauen erheben sich gegen Donald Trump“	12.12.2017
Diskurse mit nutzerseitigem Selbstregulierungseffekt (destruktive Kommentare werden durch konstruktive, lösungsorientierte Kommentare neutralisiert)	Deutschlandradio	„Thea Dorn zur Sexismus-Debatte: ‚Ein neuer Totalitarismus‘“	10.11.2017
	Deutschlandradio	„Sollen wir die AfD wie jede andere Partei behandeln? Liane Bednarz vs. Michel Friedman“	31.10.2017
	Deutschlandradio	„Die Staatsfunk-Kampagne wird weitergehen. Ein Kommentar von Brigitte Baetz“	20.10.2017
	RP Online	„Flucht aus Syrien: Wiedersehen nach 1162 Tagen“	17.7.2017
	RTL Aktuell	„GroKo wäre eine Koalition der Verlierer – aber wo sind die Alternativen“	29.12.2017
	Tagesschau.de	„Alle Jahre wieder: Gerüchte über Weihnachtsmärkte“	17.11.2017
Diskurse mit überwiegend konstruktiven, lösungsorientierten und/oder bestätigenden/bestärkenden Kommentaren (mit viel/wenig Moderationsaktivität der Redaktion)	Deutschlandradio	„Was ist Ihre Lieblingsband aus der DDR?“	16.10.2017
	RP Online	„DEG Winterwelt in Düsseldorf: Eisbahn auf der Kö wird ab nächste Woche aufgebaut“	2.11.2017
Diskurse mit lebensnahem Alltagsbezug und hohem, weitgehend neutralem Kommentaraufkommen	RP Online	„Unfall: Frau verursacht Totalschaden wegen Spinne im Auto“	5.4.2017
	Tagesschau.de	„LKW der Zukunft?“	17.11.2017

ONLINE-DISKURSANALYSE: KERNERGEBNISSE (AUSZUG)

1 Es findet **kaum redaktionelle Moderation** im Sinne aktiver Diskussionsbeiträge statt, was dazu führt, dass Redaktionen nur wenig Einfluss auf den Verlauf der öffentlichen Diskurse zu ihren Nachrichtenangeboten ausüben.

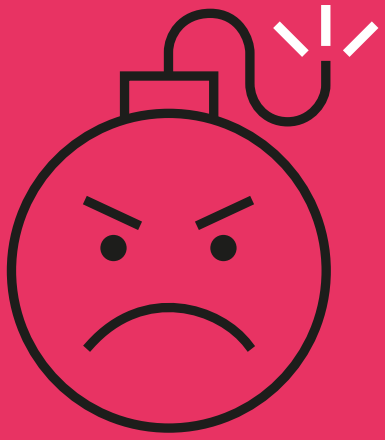
2 Eine **aktive Diskussionsbeteiligung** der jeweiligen Redaktion wirkt sich direkt auf die Positionierung des betreffenden Hauptkommentars und dessen Diskussionsstrang in der aktuellen algorithmischen **Sortierung** aus. Die entsprechenden Redaktionskommentare werden automatisch höher gerankt. Auf diese Weise kann eine Redaktion durch **gezielte Kommentierung** bestimmten Diskussionssträngen zu mehr Aufmerksamkeit verhelfen.

3 Der **Vorwurf der Propaganda und der Lügenpresse** zieht sich durch nahezu alle analysierten Diskurse. **Unabhängig vom Thema des journalistischen Beitrags** werfen Nutzerinnen und Nutzer den Journalistinnen und Journalisten bewusste Manipulation und interessengeleitete Berichterstattung vor.

4 **Höchstens ein Drittel der Kommentare hat einen thematischen Bezug** zum jeweiligen redaktionellen Beitrag. Dagegen greifen viele Kommentare Formulierungen auf, die Redaktionen in ihren Beiträgen wählen, wandeln sie ab und/oder stellen sie in neue Textzusammenhänge. Die Mehrheit dieser Kommentare besteht aus themenfremder Meinungsäußerung, häufig aus Verunglimpfungen und Hetze.

5 Es gibt nur **wenige einflussreiche, dafür jedoch durchweg negativ kommentierende Nutzerinnen und Nutzer,** die in ihren meist ähnlichen Kommentaren zu bestimmten Sichtweisen und/oder Handlungen aufrufen. Solche Art der Kommentierung scheint auf ein typisches ‚Troll'-Verhalten hinzudeuten: Diese Wiederholungstäterinnen und Wiederholungstäter sind spezielle Charaktere, deren Motive sich zwischen Geltungsdrang und missionarischem Eifer bewegen.

6 Fast alle Kommentare werden **am ersten Tag nach der Veröffentlichung** eines Beitrags erstellt, danach kommentieren Nutzerinnen und Nutzer nur noch vereinzelt und sporadisch. Ferner: Je später Kommentare im Diskursverlauf erscheinen, sind diese umso kürzer und weisen weniger Bezug zum journalistischen Beitrag auf.



REDAKTIONELLE MODERATIONS-STRATEGIEN GEGEN HASSREDE

Eine fachgerechte Kommentarmoderation kann in sozialen (und dissozialen) Diskursen im Internet das Beste aus den eingebrachten Argumenten, Einschätzungen, Positionen und Ansichten der Nutzerinnen und Nutzer herausfiltern – und das Schlechteste wegmoderieren oder ausblenden. Das ist zumeist die Erwartung all jener Redaktionen, die sich von einem professionellen Community-Management einen Mehrwert erhoffen. In der konkreten Anwendungspraxis von Journalistinnen und Journalisten stehen allerdings häufig weder genügend redaktionelle Ressourcen noch ausreichend journalistisch geschultes Personal zur Verfügung, die diese Erwartungen kompetent erfüllen könnten.

Vielfach sind Social-Media-Redakteurinnen und -Redakteure sowie deren (studentische) Aushilfen dauerhaft drastischen Stress- und Belastungsfaktoren durch explizite Inhalte in Wort und Bild (v. a. Hassrede, Rassismus, Antisemitismus, Gewaltaufrufe) ausgesetzt, die teilweise auch gegen sie persönlich gerichtet sind. Meist müssen Redaktionen in wenigen Sekunden entscheiden, ob sie Nutzerbeiträge löschen oder veröffentlichen. Das, was sie im Zeitraffer erleben, kann unter Umständen schwerwiegende psychische Folgen haben. Bis dato gibt es kaum therapeutische

Angebote seitens der Redaktionen, um diese traumatischen Erfahrungen aufzuarbeiten oder mit anderen Betroffenen zu teilen. Die Auffassung, die Entwicklung von automatisierten Filtersystemen könnte *die* lang ersehnte Lösung darstellen, um solche Inhalte zu kontrollieren und damit Diskurse durch Löschpraktiken zu zivilisieren, ist strittig.

Nachfolgend stellen wir zwei Moderationsstrategien mit insgesamt zehn Moderationselementen vor, die es Redaktionen ermöglichen sollen, die zunehmende Verbreitung von Hassrede in Nutzerdiskursen nicht nur zu bewältigen, sondern wirksam zu kontrollieren. In Zusammenarbeit mit der Rheinischen Post Online haben wir die aus der Diskursanalyse abgeleiteten Strategien in der Praxis getestet. Sie können erfolgreich gezielt gegen Hassrede eingesetzt werden.

Die quantitativ-qualitative Diskursanalyse, die eine der Grundlagen für das vorliegende Whitepaper im Auftrag der Landesanstalt für Medien NRW bildet, bezieht sich sowohl auf die Kommentarabschnitte auf den Websites der vier untersuchten Qualitätsmedien als auch auf die Kommentarbereiche von ausgewählten, redaktionell betreuten Social-Media-Präsenzen. Im Ergebnis der empirischen Untersuchung haben sich hinsichtlich der beiden Kommentarveröffentlichungsorte zwei voneinander abgrenzbare Debattenkulturen herausgebildet:

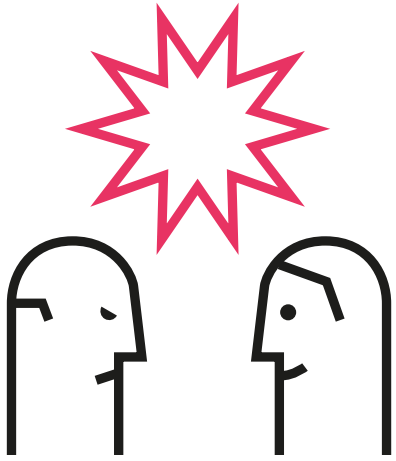
a) bei *Kommentaren auf der eigenen Website* geht es in der Moderation direkter um diskursive Inhalte, die das betreffende Nachrichtenangebot bereitstellt, oder um Feedback zur Redaktion als Organisation bzw. um einzelne Autorinnen und Autoren sowie Redakteurinnen und Redakteure oder um die Medienmarke als Ganzes;

b) bei *Kommentaren auf den untersuchten Social-Media-Kanälen* werden Diskussionen von den Nutzerinnen und Nutzern dominiert, d. h. die Redaktion nimmt eine untergeordnete Moderationsrolle ein, weil die ausgelagerten Debatten einen allgemeineren lebensnahen Bezug haben und sich Nutzerinnen und Nutzer stärker aufeinander beziehen.

Dass sich hierbei auch die deliberative Qualität von Kommentaren teilweise unterscheidet (auf Social-Media-Kanälen war sie geringer) und ein direkter Zusammenhang zwischen Teilnahmehürden und der Zivilität von Nutzerdebatten besteht, liegt vor allem in der Natur der jeweiligen

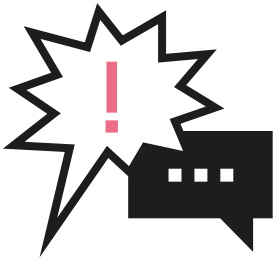
Kommentarinfrastruktur und ihrer Technologie: Auf den Social-Media-Kanälen des Mediums posten Nutzerinnen und Nutzer in der Regel Beiträge von ihren eigenen Accounts und oftmals im Gefolge der persönlichen Netzwerke (und erhalten dadurch aufgrund der Algorithmisierungslogik mehr Feedback von „Freundinnen und Freunden" zu diesen Beiträgen). Dagegen müssen sie sich auf den Redaktions-Websites häufig aufwändigen Registrierungsverfahren oder einer strengen Überprüfung ihres Profils seitens der Redaktion unterwerfen. Auch machen Redaktionen auf ihren Social-Media-Kanälen weniger von ihrem „Hausrecht" Gebrauch und achten bei Kommentierenden in der Regel seltener auf die Einhaltung ihrer Netiquette. Dass dies vor allem eine Ressourcenfrage ist, hat sich in den 12 Expertengesprächen immer wieder bestätigt. Überraschenderweise resignieren viele Redaktionen alleine schon aufgrund der Kommentarmenge, die ihrer Aussage zufolge nicht mehr handhabbar sei.

Im *experimentellen Live-Betrieb bei RP Online* wurde deutlich: Vor allem die Grundsätze der *Ermächtigung und Entmachtung* (Empowerment/Dis-Empowerment) von Kommentierenden, kompletten Diskursen oder einzelnen Diskurselementen sollten als praxistaugliche Moderationsstrategien zur Anwendung kommen. Psychologinnen und Psychologen haben herausgefunden, dass ein System von Belohnung und Bestrafung im Umgang mit Haterinnen und Hatern durchaus erfolgreich sein kann: Feldversuche zeigen, dass solche lernpsychologischen Anreize für



soziales bzw. gegen antisoziales Verhalten in Kommentarsektionen Wirkung zeigen, wenn man Haterinnen und Hater getreu der Devise *„Don't Feed the Trolls!"* die Bühnen der öffentlichen Verunglimpfung entzieht. In der Gesamtbetrachtung können wir in Bezug auf die Kommentarmoderation dahingehend mindestens drei idealtypische Ansätze zur Begrenzung von Hassrede und zur Anregung von konstruktiven Diskursen benennen, die sich übergreifend in eine *protektive* (Nutzerinnen und Nutzer schützende), eine *disqualifizierende* (Nutzerinnen und Nutzer ausschließende) und eine *unterstützende* Herangehensweise unterteilen lassen. Dabei verstärkt die Bandbreite provokanter Herangehensweisen – zum Beispiel ein übermäßig raues, durch die Moderation provoziertes Kommunikationsklima – das Unwesen von Haterinnen und Hatern sowie Trollen nachweislich.

Entsprechend haben wir uns im Test der Moderationsstrategien zur Zivilisierung von Netzdiskursen vor allem auf das Protektionismus-, Unterstützungs- und das Exklusionsprinzip von Nutzerinnen und Nutzern konzentriert. Doch wurden im Praxistest zeitweise auch provokante Elemente in Form von Ironie getestet. Insgesamt lassen sich, je nach Intensität und Aktionsradius, zehn verschiedene Moderationselemente auf einem Kreuzdiagramm eintragen, das dem Vierfeldschema starkes/schwaches „Dis-Empowerment" vs. starkes/schwaches „Empowerment" und (eher) starker vs. (eher) schwacher „redaktioneller Aktivitätsgrad" folgt (Abb. 2).



REGULIERENDE STRATEGIE: DIS-EMPOWERMENT

Punishment (Bestrafung)

Straf-/Zivilrechtliche Konsequenzen gegen Hassrede

Diese Herangehensweise ist Teil einer Ausschlussstrategie, wenn Kommentarmoderatorinnen und Kommentarmoderatoren gefordert sind, Nutzerinnen und Nutzer abzumahnen oder bei entsprechenden Meinungsäußerungen strafrechtlich verfolgen zu lassen. Punishment (Bestrafung) zeichnet auf der y-Achse ein eher mittlerer redaktioneller Aktivitätsgrad, auf der x-Achse aber die höchste Form des Dis-Empowerments aus: Verleumdungen, Beleidigungen und Volksverhetzung sind ebenso wenig von der Meinungsfreiheit gedeckt, wie Aufrufe zu Straf- oder Gewalttaten keine Kavaliersdelikte sind. Es sind vielmehr aktive Tatbestände, die strafrechtlich verfolgt und mit empfindlichen Freiheitsstrafen geahndet werden können – das gilt nicht nur für Beiträge auf der eigenen Website des Medienunternehmens, sondern auch für deren Social-Media-Kanäle: Einer/Einem möglicherweise über die IP-Adresse aufgespürten Täterin/Täter drohen strafrechtliche Sanktionen wie eine Freiheits- oder eine Geldstrafe nach §§ 185, 186, 187 StGB.

Counter Speech (Gegenrede)

Argumentatives Sprechen gegen Hassrede

Im Moderationsprozess ist diese Strategie eine der aufwändigsten, weil sie eine aktive Kommunikation und vor allem eine hohe Aufmerksamkeit sowie dynamisches Handeln seitens der Kommentarmoderation erfordert. Sie ist aber auch der wichtigste Versuch, um dissoziale Diskurse in konstruktive Dialoge zu verwandeln, ohne einzelne Nutzerinnen und Nutzer sperren oder deren Beiträge löschen zu lassen: Dementsprechend ist die Gegenrede – damit sind ebenso Gegenpöbeleien der Moderatorinnen und Moderatoren gemeint, auch wenn wir diese nicht empfehlen – durch einen vergleichsweise hohen redaktionellen Aktivitätsgrad und starkes Dis-Empowerment definiert. Argumentatives Sprechen gegen Hassrede ist – sowohl von Nutzer- als auch von Redaktionsseite – zudem von der Haltung geprägt, sich nicht alles gefallen zu lassen, was von Haterinnen und Hatern gepostet wird. Die elegante Variante für ressourcenschwache Redaktionen ist es, die Gegenrednerinnen und Gegenredner unter den Nutzerinnen und Nutzern durch aufmunternde Kommentare zu belohnen (vgl. „Embracing"), statt selbst andauernd Gegenrede zu betreiben. Das stärkt die loyalen Nutzerkreise und wirkt sich positiv auf eine selbstregulierende Debattenkultur aus.

Deconstructing (Zerlegen)

Hassrednerinnen und Hassredner sowie Hassrede dekonstruieren

Hassrednerinnen und Hassredner zu dekonstruieren kann entmutigend sein und ist eine Strategie, die auch Redaktionen mit hohem Personalschlüssel kaum zu leisten imstande sind, weil sie eine Art Chefarztbehandlung für Haterinnen und Hater sowie Trolle bedeutet, für die wohl kaum ein Medienhaus entsprechend viel Zeit und Geld investieren kann. Einerseits. Andererseits erscheint es sinnvoll, gerade Meinungsäußerungen dieser lauten Minderheit bis ins Detail zu widerlegen, damit sie die schweigende Mehrheit z. B. nicht mit populistischen Falschaussagen beeinflusst. Der Aufwand, den ein hoher redaktioneller Aktivitätsgrad bei dieser Herangehensweise bedeutet, erscheint dann gerechtfertigt, wenn das Dekonstruktionsprinzip bei Hassrednerinnen und Hassrednern anschlägt und nachhaltige Wirkung für die gesamte Diskussionskultur des jeweiligen Mediums zeigt. Das Dis-Empowerment ist hier etwas schwächer ausgeprägt, da Moderatorinnen und Moderatoren sich auf das Niveau von Haterinnen und Hatern „einlassen" und sich in ihre Argumentationsketten hineindenken müssen.

Blocking/Deleting (Ausblenden)

Hassrednerinnen und Hassredner stummschalten, Hassredebeiträge löschen/ausblenden

Diese Strategie ist in deutschen Medienhäusern derzeit eine der gängigsten, weil zwar am preisgünstigsten und unaufwändigsten – aber auch eine der am wenigsten nachhaltigen: Ob sie intern von Redaktionen im Schichtbetrieb erledigt oder an einen Dienstleister ausgelagert wird: Die Löschpraxis von Hassrede hat sich gerade bei Kommentaren durchgesetzt. Allerdings ohne bleibenden Erfolg: Viele Hassrednerinnen und Hassredner sowie Störerinnen und Störer kehren mit neu aufgesetzten Social-Media-Profilen unter anderem Namen immer wieder zurück und machen genauso weiter wie zuvor. Als effiziente Variante hat sich die folgende Herangehensweise herausgestellt: Hassrede nicht löschen, sondern (durch redaktionelle Zusatzfunktionen) deren Urheberinnen und Urheber stummschalten und ihre Kommentare für andere Nutzerinnen und Nutzer unsichtbar machen, während diese für sie selbst noch sichtbar bleiben. Auch hierbei bleibt die redaktionelle Aktivität vergleichsweise gering und das Dis-Empowerment ist relativ hoch.

Ignorance (Aufmerksamkeitsentzug)

Keine Reaktion auf Hassrede, ignorieren

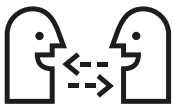
Negativkommentare komplett zu ignorieren kann ebenfalls ein Moderationselement sein, wenn auch ein eher passives, das ein Minimum an redaktionellem Aufwand bei gleichzeitig begrenztem Dis-Empowerment bedeutet. Das Problem für die Redaktionen ist: Sollte sich in den redaktionell betreuten Kommentarbereichen eine Nutzerdiskussion entfalten, die möglicherweise strafrechtliche Tatbestände enthält, kann die Redaktion am Ende unter Umständen dafür (mit-)verantwortlich gemacht werden. Statt Hassrede ganz zu ignorieren, bietet es sich für die Moderation eher an, einzelne Störenfriede bewusst zu übersehen und nicht auf ihre Hassbeiträge zu reagieren, um ihren Argumenten die Aufmerksamkeit zu entziehen. Dieser Aufmerksamkeitsentzug ist eine mutige, sehr grundsätzliche redaktionelle Entscheidung, weil sich die Eigendynamik von Diskursen im Verantwortungsbereich der Redaktion gerade bei Social Media damit extrem verschärfen kann.

BESTÄRKENDE STRATEGIE: EMPOWERMENT

Ironization (Ironie/Humor)

Hassrede ironisieren und mit Humor begegnen

Zwar stellt es keine wünschenswerte Moderationsstrategie dar, Hassrede zu ironisieren, dennoch kommt diese Methode häufig zur Anwendung: In vielen Redaktionen begegnen Moderatorinnen und Moderatoren den Negativkommentaren von Nutzerinnen und Nutzern mit Ironie, inzwischen auch mit Sarkasmus und Zynismus – teils aus Frustration, teils aus Hilflosigkeit. Im ironischen Unterton und mit ironisierender Sprache, unter Einsatz von GIFs, Emojis oder Likes versuchen sie, dissoziale Diskurse ins Positive zu wenden, was jedoch mitunter kontraproduktiv sein kann: Im schriftlichen Austausch stößt Ironie oft auf Unverständnis, umso mehr, wenn sich Diskutierende nicht persönlich kennen. Auch wenn der vergleichsweise geringe redaktionelle Aktivitätsgrad für derlei Moderationselemente dazu verleitet, gilt die Regel im Journalismus auch generell für die schriftliche Kommentarmoderation: Ironie sollte, wenn überhaupt, von Moderatorinnen und Moderatoren nur selten eingesetzt und nicht dazu verwendet werden, Störerinnen und Störer sowie Haterinnen und Hater zu verunglimpfen oder zu verspotten, weil es im Zweifel negativ auf die Redaktion zurückfallen kann, zumal sich hierbei das Empowerment in Grenzen hält.



Understanding (Verständnis zeigen)

Hintergründe von Hassrede ermitteln, Versachlichung der Debatte

Ein weiteres Strategieelement, das einen recht hohen redaktionellen Aktivitätsgrad und damit Ressourcen seitens der Redaktionen voraussetzt, ist, Verständnis gegenüber Hassrednern, Pöblern und Störern zu zeigen. Im Spektrum des Empowerments ist dieser vor allem als Argumentetransfer gemeinte Ansatz nicht so zeitaufwändig wie ein intensiver Dialog, aber schon das gezielte Nachfragen nach den Hintergründen von hasserfüllten Kommentaren, den Beweggründen der Urheberinnen und Urheber sowie die Einschätzung der zugrundeliegenden (moralischen) Normen mit einer gründlich moderierten Wertediskussion erfordern nicht nur Beharrlichkeit, sondern auch stabile Nerven und geschultes Personal. Dabei kann es sich durchaus lohnen, eine vergiftete Diskussion ausdauernd zu versachlichen und im Kontext von Hassrede Klarheit zu schaffen, indem auf den Austausch von Argumenten und persönliche Begründungen gedrungen wird. Das zeigt sich nicht zuletzt in der nach außen vertretenen Haltung der jeweiligen Medienmarke, die sich durch eine solche Strategie redaktionell positionieren und abgrenzen kann.

Dialogization (Vermittelnder Dialog)

Zwischen Gegenpositionen vermitteln, Dialog zwischen Nutzerinnen und Nutzern anregen

Die Königsdisziplin unter den Strategieelementen heißt Dialogisierung: Es geht dabei z. B. um die Vermittlung von Gegenpositionen, die mitunter auch Hassrede enthalten können. Anders als beim „Understanding" versuchen Moderatorinnen und Moderatoren nicht nur korrigierend, sondern auch vermittelnd einzugreifen, um einen Dialog anzuregen, der sich nicht in emotionalen Scharmützeln verliert. Die Hoffnung ist, dass sich auch Negativkommentatorinnen und Negativkommentatoren für einen konstruktiven Dialog gewinnen lassen, wenn sie sich auf eine Sachargumentationsebene begeben. Auf Moderationsseite bedingt das, sich auch extremen Positionen gegenüber zu öffnen, selbst wenn diese auf falschen Tatsachen aufbauen oder Beleidigungen enthalten. Weil das Abwägen solcher Positionen sehr nervenaufreibend sein kann und der redaktionelle Aktivitätsgrad äußerst hoch ist, empfiehlt sich in der Regel eine Doppelbesetzung von Social-Media-Moderationsschichten. Mit Enttäuschungen in alle Richtungen muss man rechnen – allerdings mit der Aussicht darauf, dass auch Haterinnen und Hater letztlich nur Menschen sind und sich irgendwann bei ihrer Ehre gepackt sehen, zivil mitzudiskutieren. Ist das erreicht, wurde dieses Strategieelement erfolgreich umgesetzt und das hohe Empowerment-Potenzial entsprechend ausgereizt.

Solidarization (Gemein machen)

Sich mit Betroffenen und Gegnerinnen und Gegnern von Hassrede solidarisieren

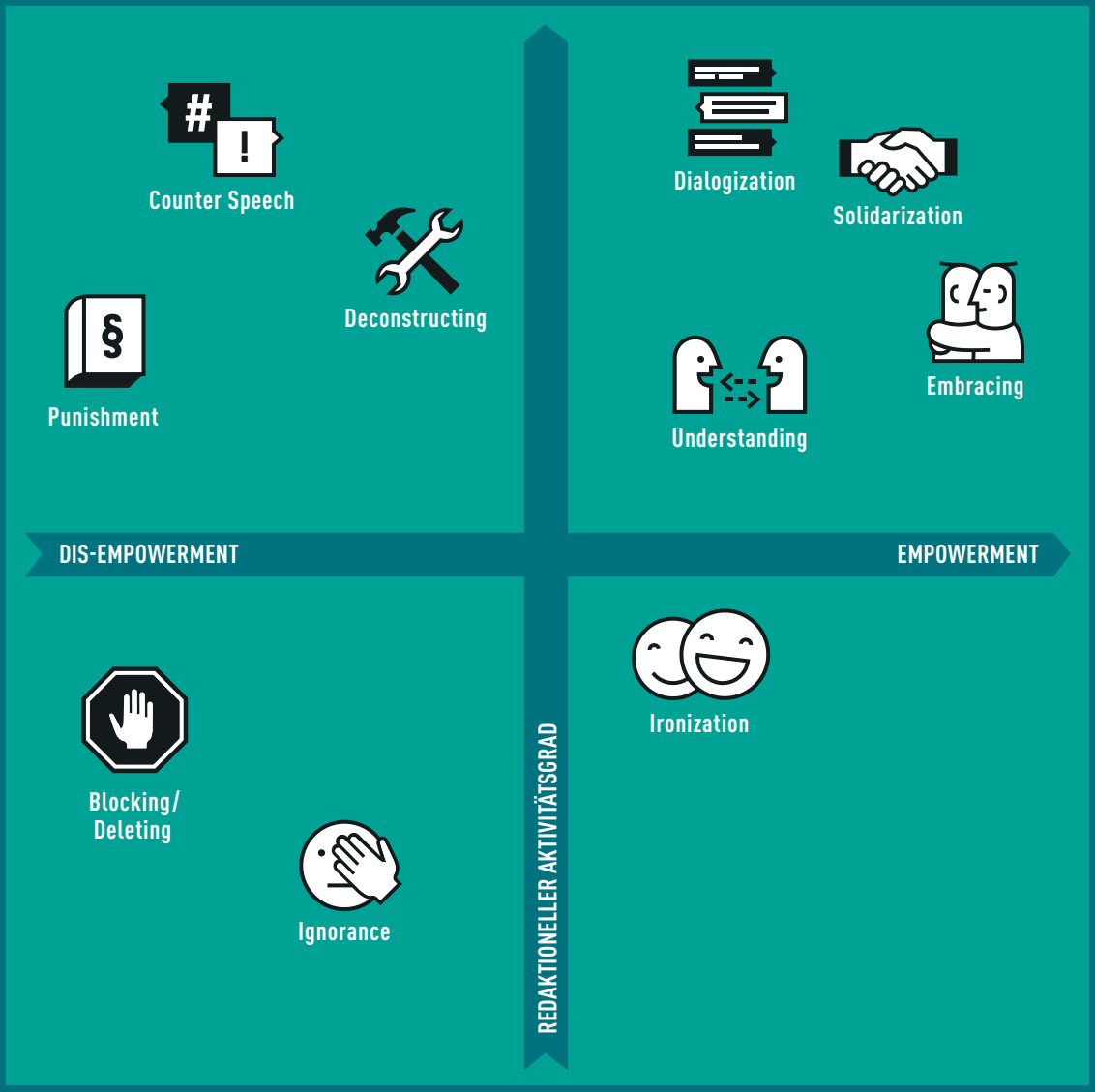
Bei diesem Moderationselement werden die Grenzen journalistischer Professionalität teilweise überschritten. Trotzdem soll es hier wegen seines strategischen Einflusses positive Beachtung finden: Sich mit einer Sache gemein machen, auch mit einer guten, gilt im Journalismus spätestens seit Hanns Joachim Friedrichs als unschicklich, kann aber in Nutzerdiskursen interessante Effekte mit sich bringen. Zum einen erweist sich diese Methode als nützlich, weil sich Kommentarbereiche plötzlich zu hassredefreien Zonen entwickeln können, indem Haterinnen und Hater sowie Störerinnen und Störer aufgrund der hohen Solidarität der Redaktion mit ihren Gegnerinnen und Gegnern (sowie Gegenrednerinnen und Gegenrednern) dort nicht mehr geduldet werden. Zum anderen formiert sich die Gruppe der Gegnerinnen und Gegnern von Hassrede möglicherweise zu einer höchstloyalen Nutzerschaft, die sich nicht nur überzeugter in Diskussionen einbringt, sondern sich auch stärker mit der Medienmarke identifiziert und weitere Nutzerinnen und Nutzer an diese bindet. Anders gesagt: Solidarisierungskonzepte wie in der #metoo-Debatte können sich zu kampagnenhaften Solidarisierungswellen auswachsen, die auf das Medium und seine journalistische Haltung einzahlen. Der Strategie-Matrix folgend (Abb. 2) hängt diese Methode mit einem hohen Empowerment zusammen, das sich durch eine entsprechend hohe redaktionelle Aktivität steigern lässt.

Embracing („Umarmen", motivieren)

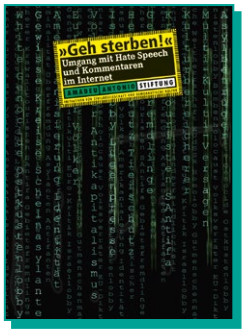
Betroffene sowie Gegenrednerinnen und Gegenredner zielgerichtet im Diskurs stärken

Die Umarmungs-Methode bezieht sich auf Gegenrednerinnen und Gegenredner sowie Zielsubjekte von Hassrede, indem diese in ihrer Argumentation gezielt gestützt und damit in ihrer Kommunikation gegen und gegenüber Hassrednerinnen und Hassrednern gestärkt werden. Anders als bei der offenen Gegenrede wird hierbei jedoch auf explizite Sprache, Ermahnungen gegen Hassrednerinnen und Hassredner, Stimmungsmache und dergleichen verzichtet; stattdessen wird die Gegenseite verbal bekräftigt und durch die Präsenz der Moderation moralisch unterstützt. Auch hier wird ein hoher redaktioneller Aktivitätsgrad vorausgesetzt, damit das Empowerment funktionieren kann. Haterinnen und Hater sowie Störerinnen und Störer werden bei dieser Herangehensweise weder ignoriert, noch ausgeblendet oder bestraft. Der Diskurs wird dadurch verbessert, dass sie sich durch ihre fragwürdige Argumentation im besten Fall selbst disqualifizieren, wodurch sich die Angegriffenen besser geschützt fühlen. Eine Kombination mit regulierenden Moderationselementen wie z. B. „Counter Speech" oder einem zusätzlichen Empowerment durch „Solidarization" erscheint sinnvoll.

Abbildung 2: Redaktionelle Moderationsstrategien gegen Hassrede – Matrix regulierender und motivierender Moderationselemente



ANHANG



LEITFÄDEN UND AUTOMATISIERUNGS-TOOLS FÜR REDAKTIONEN IM UMGANG MIT HASSREDE

Leitfäden

NO HATE SPEECH MOVEMENT:
Leitfaden für Journalistinnen und Journalisten im Umgang mit Hate Speech im Netz (2017)

Der Leitfaden des „No Hate Speech Movement", koordiniert von der gemeinnützigen Organisation Neue Deutsche Medienmacher, enthält hauptsächlich konkrete Handlungsempfehlungen für Nachrichtenredaktionen und individuelle Journalistinnen und Journalisten im Umgang mit Hassrede im Internet. Ziel ist es, Hinweise und Empfehlungen bereitzustellen, wie mit Hassrede in sozialen Netzwerken, vor allem bei Facebook, umgegangen werden kann – sowohl bei Hassrede, die Redaktionen und journalistische Angebote persönlich angreift, als auch bei solcher ohne direkten Bezug zu Journalistinnen und Journalisten als Personen oder Organisation.

→ Download als PDF

AMADEU ANTONIO STIFTUNG:
„Geh sterben!“ Umgang mit Hate Speech und Kommentaren im Internet (2016)

Die von der Amadeu Antonio Stiftung herausgegebene Broschüre thematisiert Hassrede als Phänomen in einer zunehmend von Online-Kommunikation geprägten Gesellschaft. Die Publikation hat zum Ziel, einen Überblick über die Debatte zum Thema Hate Speech zu geben und Lösungsansätze zu formulieren. In diesem Rahmen liefert sie eine definitorische Einordnung, listet Merkmale von Hasskommunikation auf, um diese schneller erkennen zu können, und ermöglicht dank Erfahrungsberichten von Journalistinnen und Journalisten sowie Opfern sogenannter Shitstorms einen Einblick, wie Betroffene Hassrede wahrnehmen, die gegen ihre Person oder ihre Redaktionen gerichtet ist.
→ Download als PDF

WAN-IFRA:
Do Comments Matter?
Global Online Commenting Study (2016)

Die World Association of Newspapers and News Publishers beschäftigt sich in ihrer Studie „Do Comments Matter?“ mit dem Mehrwert von Kommentaren zu Online-Nachrichten im redaktionellen Kontext und untersucht vorrangig, wie sich Journalistinnen und Journalisten weltweit gegenüber Nutzerdiskursen verhalten. Ziel der empirischen Untersuchung ist es, Handreichungen für die redaktionelle Moderation von Kommentaren zu geben und Beispiele aufzuzeigen, in denen es Medienorganisationen gelingt, konstruktive Gespräche mit ihren Zielgruppen zu fördern. Dazu wurden weltweit 78 Medienunternehmen aus 46 Ländern in persönlichen Interviews und via Online-Umfrage zu ihren Eindrücken und ihrem praktischen Umgang mit Nutzerkommentaren befragt.
→ Download als PDF

THE CORAL PROJECT:
Community Guides for Journalism. Instructions and ideas for better engagement, written by experts (2017)

Das Coral Project ist eine Verbundinitiative unter Beteiligung der Mozilla Foundation, der New York Times und der Washington Post, die die Qualität insbesondere von journalistischen Online-Diskussionen zu steigern beabsichtigt. Im Rahmen des Projektes entstehen Open-Source-Werkzeuge, die von Redaktionen kostenfrei für das Management ihrer Communities eingesetzt werden können. In der „Guide“-Rubrik auf der Website des Coral Projects findet sich ein umfangreicher Leitfaden für Nachrichtenredaktionen zur Online-Interaktion mit ihren Communities. Der Leitfaden verfolgt das Ziel, Redaktionen beim Aufbau und der Pflege ihrer Communities zu helfen. Er enthält rund 70 verschiedene Managementstrategien, Fallstudien von Medienschaffenden aus der ganzen Welt und weiterführende Quellen, die den Prozess eines erfolgreichen Umgangs mit der Online-Leserschaft beschreiben, von Zielgruppenstrategien bis hin zur Verwendung von Analyse-Tools.
→ Download als PDF

Automatisierungstools

CONVERSARIO:
„Mehr Zeit für guten Dialog“

Weil die Kommentarmenge und die Dynamik von Diskursen im Netz manuell kaum beherrschbar sind, verspricht Conversario „proaktiven Schutz gegen Hass- und Spam-Kommentare“. Als erstes deutsches Startup konzentriert sich Conversario auf KI-basierte Kommentarmoderation in den sozialen Medien. Hinter Conversario steht das Technologieunternehmen ferret go GmbH mit Sitz in Bernau bei Berlin, das sich schwerpunktmäßig mit Natural Language Processing, Machine Learning und Automated Services befasst. Mit Conversario hat die Firma ein Tool für automatisiertes Community Management entwickelt, das den Nutzerdialog positiv beeinflussen will. Ferret go arbeitet laut Website bereits mit namhaften deutschen Verlagen und Medienhäusern zusammen, darunter Focus Online, FAZ.net, n-tv, der Berliner Zeitung und dem rbb. Auch bei Tagesschau.de wird die Implementierung von Conversario in Erwägung gezogen, um unbedenkliche Kommentare bei Facebook und auf der eigenen Website zielsicher herauszufiltern.
→ https://conversar.io/de

TALK:
„Have better conversations“

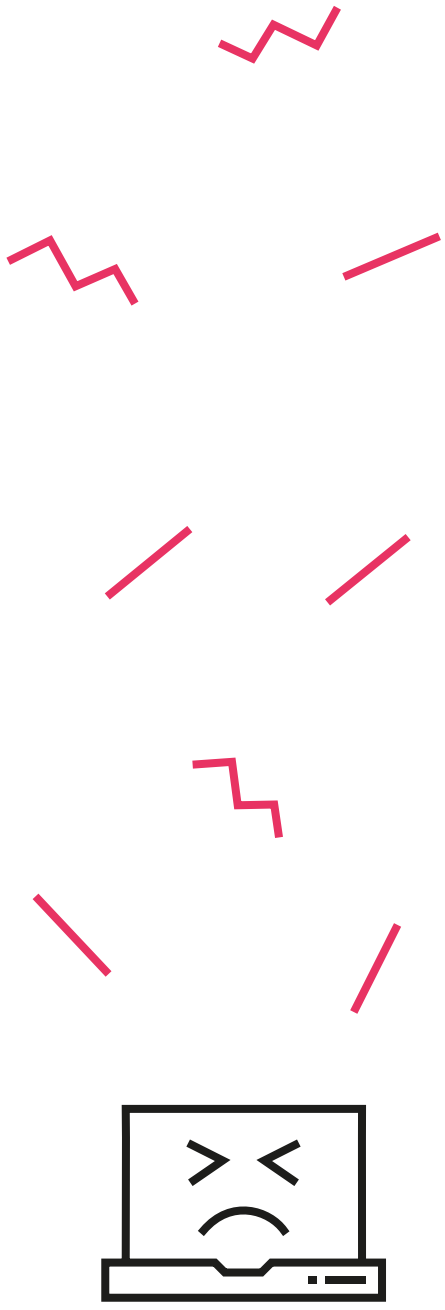
Die von der gemeinnützigen John S. and James L. Knight Foundation finanzierte Software „Talk“ wurde ursprünglich in Kooperation mit Washington Post und New York Times maßgeblich vom Coral Project entwickelt, von dem sie auch betrieben und vermarktet wird. Das Tool will Journalistinnen und Journalisten sowie ihre Communities enger zusammenbringen, um auf diese Weise Online-Diskurse konstruktiver zu gestalten und generell das Diskussionsklima im Hinblick auf redaktionelles Audience Engagement zu verbessern. „Viele unserer treuesten Leser sind Kommentatoren. Die Kombination von Talk und ModBot (eine von der Post entwickelte Moderationssoftware, die auf Künstlicher Intelligenz beruht) ermöglicht es uns, diese besser kennenzulernen, leichter mit ihnen zu interagieren und schnell nachdenkliche und aufschlussreiche Kommentare für alle Leser zu finden und hervorzuheben“, erklärte Emilio Garcia-Ruiz, Chefredakteur der Post, zur Einführung des Tools. „Dies ist ein einzigartiges Kommentierungssystem, das einen umfassenden Ansatz für Kommentare bietet und uns die technischen Möglichkeiten aufzeigt, sich mit Kommentatoren auf eine tiefere und sinnvollere Weise zu befassen.“
→ https://www.coralproject.net/talk

PERSPECTIVE API:

„What if technology could help improve conversations online?“

Bislang ist es noch keinem Software-Anbieter gelungen, ein intelligentes System zu entwickeln, das die Hygiene von Netzdiskursen nennenswert verbessert und durch den Einsatz von Algorithmen Diskriminierung verhindern hilft. Das Google-Schwesterunternehmen Jigsaw – Vorläufer: „Google Ideas“ (2010–2015) – hat als Tech-Inkubator mit Sitz in New York City eine ambitionierte API (Application Programming Interface) entwickelt, die Textkommentare auf einer „Toxizitäts“-Skala von 0 bis 100 bewertet, um Nachrichtenanbietern die Vormoderation und Bewertung von Kommentaren zu erleichtern. Zu den ersten Kooperationspartnern gehören namhafte Medienhäuser wie die New York Times, der Economist, der Guardian und die Wikimedia Foundation, Betreiberin der Online-Enzyklopädie Wikipedia, die zunächst 115.000 Diskussionsbeiträge für den ersten maschinellen Lernprozess von Perspective bereitstellen. Das Tool soll erkennen können, ob Nutzerinnen und Nutzer einen Kommentar als unangemessen bewerten würden und, darüber hinaus, unter welchen widrigen Bedingungen sie eine Diskussionsplattform verlassen würden. Perspective API sortiert und klassifiziert automatisch, die News-Betreiber entscheiden jedoch individuell (und durch menschliche Moderatorinnen und Moderatoren gesteuert), ob die als toxisch eingestuften Kommentare gelöscht, ausgeblendet, strafverfolgt, veröffentlicht oder einem anderen Zweck zugeführt werden sollen.

→ https://www.perspectiveapi.com



DIE AUTOREN

Foto: Beate C. Koehler

Dr. phil Leif Kramp ist Senior Researcher und Forschungskoordinator am Zentrum für Medien-, Kommunikations- und Informationsforschung (ZeMKI) der Universität Bremen. Kramp gehört zum Gründungsvorstand des Vereins für Medien- und Journalismuskritik e.V., ist Mitglied des Direktoriums des journalistischen Nachwuchsförderprogramms VOCER Innovation Medialab sowie Autor und Mit-Herausgeber zahlreicher Fachbücher und Studien zur Transformation des Journalismus. Er ist Mitglied der Nominierungskommission des Grimme Online Awards 2018, der Jury des NETZWENDE Awards für nachhaltige Innovationen im Journalismus und der Jury der Initiative Nachrichtenaufklärung e.V.

Kontakt: kramp@uni-bremen.de

Foto: Jörg Möller

Prof. Dr. phil. Stephan Weichert leitet den Masterstudiengang „Digital Journalism", das „Urban Storytelling Lab" und das Digital-Journalism-Fellowship-Programm an der Hamburg Media School (HMS). Seit 2008 lehrt er als Professor für Journalistik in Hamburg. Weichert ist Gründer des Think Tanks VOCER.org und Gründungsdirektor des VOCER Innovation Medialab, eines Stipendienprogramms für Nachwuchsjournalisten. Seit 20 Jahren setzt er sich als Wissenschaftler und Publizist mit den Folgen der Digitalisierung für Medien, Journalismus und Gesellschaft auseinander. Für seine herausragende journalistische Arbeit zur „Digitalen Gesellschaft" wurde Weichert im Jahr 2014 der „Medienethik-Award" verliehen.

Kontakt: stephanweichertpost@gmail.com

Zuletzt veröffentlichten Leif Kramp und Stephan Weichert eine umfangreiche Studiensammlung zur Mediennutzung junger Zielgruppen, u.a. ist bei VISTAS ihr Buch „Der Millennial Code. Junge Mediennutzer verstehen – und handeln" (2017) erschienen.

IMPRESSUM

Herausgeber:
Landesanstalt für Medien NRW
Zollhof 2
D-40221 Düsseldorf
T +49 211 77007-0
F +49 211 727170
info@medienanstalt-nrw.de
www.medienanstalt-nrw.de

Verantwortlich:
Sabrina Nennstiel (Leitung Kommunikation)

Redaktion:
Dr. Meike Isenberg (Forschung),
Marie-Franca Hesse (Kommunikation)

Gestaltung:
ressourcenmangel an der Panke GmbH, Berlin

Druck:
Börje Halm, Wuppertal

Mit Unterstützung der Google Germany GmbH

Google