

Care-Arbeit für die Kommentarspalten. Was Community-Manager*innen für Online-Debatten tun und aushalten

Alina Darmstadt, Antonia Graf, Oliver Saal & Teresa Sündermann
(Amadeu Antonio Stiftung | Civic.net – Aktiv gegen Hass im Netz)

Community-Management ist oft unsichtbar – aber wer verstehen will, welche Bedeutung es für die sozialen Netzwerke hat, der stelle sich kurz vor, alle Kommentarspalten blieben unmoderiert. Der vorliegende Beitrag grenzt Community-Management, das von zivilgesellschaftlichen Organisationen in Social Media geleistet wird, von anderen Formen der Moderation wie Content-Moderation oder der ehrenamtlichen Betreuung von Gruppen, Foren und Chats ab. Wir zeigen auf, welche Handlungsstrategien Community-Manager*innen zur Verbesserung von Online-Debatten und zum Schutz diskriminierter Gruppen zur Verfügung stehen. Wir diskutieren die vielfältigen Facetten dieser anspruchsvollen Tätigkeit. Wir problematisieren, dass die Arbeit von Community-Manager*innen sowohl in Kommentarspalten als auch gesellschaftlich sowie in Organisationskontexten häufig unsichtbar bleibt. Wir zeigen Wege auf, mit denen ihre Rolle und ihr gesellschaftliches Prestige gestärkt und gesteigert werden können.

Empfohlene Zitierung:

Darmstadt, Alina/Graf, Antonia/Saal, Oliver/Sündermann, Teresa (2023). Care-Arbeit für die Kommentarspalten. Was Community-Manager*innen für Online-Debatten tun und aushalten. In: Institut für Demokratie und Zivilgesellschaft (Hg.). Wissen schafft Demokratie. Schwerpunkt Netzkulturen und Plattformpolitiken, Band 14. Jena, 124–137.

Schlagwörter:

Community-Management, Seitenmoderation, Social Media, Hate Speech, Care-Arbeit



**„COMMUNITY-MANAGEMENT
KOMMT EINE ENTSCHEIDENDE
ROLLE FÜR DEN
DEMOKRATISCHEN DISKURS
AUF ONLINE-PLATTFORMEN
ZU.“**

Alina Darmstadt, Antonia Graf, Oliver Saal & Teresa Sündermann

Community-Management ist die Care-Arbeit des Internets. Wie andere Care-Arbeit auch ist die Arbeit meistens unsichtbar und wird oft schlecht, manchmal sogar gar nicht bezahlt. Mit unzureichenden Ressourcen ausgestattet, kennt sie keine Wochenenden oder Feiertage, sondern muss rund um die Uhr erledigt werden. Solange sie gewährleistet ist, wird sie als Selbstverständlichkeit hingenommen (Winkler 2015, 15). Wer verstehen will, welche Bedeutung Moderation für die sozialen Netzwerke hat, der stelle sich kurz vor, es würde sie nicht geben. Je nach Perspektive wären die meisten Social-Media-Seiten dann ein Ort zum Fürchten – oder noch mehr, als sie das sowieso schon sind.

Zuletzt erhielten Content-Moderation-Teams der großen Social-Media-Plattformen, die rund um die Uhr gewalttätige, diskriminierende und terroristische Inhalte sichten und moderieren, erhöhte Aufmerksamkeit (Gillespie 2018). Hierzulande säubern mindestens 5.000 Content-Moderator*innen für geringe Bezahlung die sozialen Netzwerke von solchen Inhalten. Die „Putzkräfte der Demokratie“ (Kloiber 2023a) sind oft gut ausgebildet und mehrsprachig, viele sind jedoch aufgrund ihres Zuwanderungsstatus auf die Arbeit angewiesen. Eine gewerkschaftliche

Überall dort, wo es um Austausch geht, sind die sozialen Netzwerke auf das Engagement von Content-Moderator*innen angewiesen, denn es gibt Milliarden Postings und Kommentare, die diskriminierend oder gewaltvoll sind und nicht durch die plattformeigene Content-Moderation bearbeitet werden.

Organisierung hat angesichts der schlechten Arbeitsbedingungen inzwischen begonnen (Welzel 2023 und Kloiber 2023b). Weltweit sind es Hunderttausende, die im Auftrag von YouTube, TikTok, Instagram und Co. für die notwendige, aber mental belastende Arbeit ihre psychische Gesundheit gefährden (Newton 2019). Allerdings halten noch weit mehr Menschen die Diskussionskultur in den Kommentarspalten aufrecht – beispielsweise auf der eigenen Instagram-Seite, als Social-Media- oder Community-Manager*in für die Seiten eines Unternehmens oder als Administrator*in der Kiezgruppe. Überall dort, wo es um Austausch geht, sind die sozialen Netzwerke auf dieses Engagement angewiesen, denn es gibt Milliarden Postings und Kommentare, die diskriminierend oder gewaltvoll sind und nicht durch die plattformeigene Content-Moderation bearbeitet werden.

Entsprechend argumentieren wir in diesem Beitrag, dass Community-Management eine entscheidende Rolle für den demokratischen Diskurs auf Online-Plattformen zukommt. Dafür erläutern wir zunächst, welchen Gesetzgebungen die Inhaltsmoderation digitaler Plattformen unterliegt und welche regulierenden und diskursfördernden Instrumente Community-Manager*innen zur Verfügung stehen. Danach problematisieren wir die doppelte Unsichtbarkeit der geleisteten Arbeit und des Engagements von Community-Manager*innen. Wir zeigen Wege auf, mit denen ihre Rolle und ihr Prestige sowohl auf der gesellschaftlichen Ebene wie auch der Organisationsebene gestärkt werden können.

Die (un)sichtbare Care-Arbeit des Community-Managements

Social-Media-Plattformen und ihre Inhalte werden durch drei unterschiedliche Formen von Inhaltsmoderation reguliert: Digitale Plattformen wie Facebook unterliegen erstens sowohl nationalen als auch EU-weiten Gesetzgebungen wie dem Digital Services Act, die sie mehr oder weniger konsequent umsetzen. Zweitens geben sich die Plattformen eigene Nutzungsbedingungen (Community-Guidelines), die definieren, welche Inhalte unerwünscht sind und deshalb durch KI-gestützte Anwendungen sowie Content-Moderator*innen gelöscht werden sollen. Im vorliegenden Beitrag beschäftigen wir uns mit einer dritten Form der Inhaltsmoderation: dem Community-Management. Im Unterschied zur plattformseitigen Content-Moderation liegt die Moderation der Kommentarspalten von Medienunternehmen, NGOs, Behörden und Agenturen, des Kommentarbereichs auf Nachrichtenseiten oder die Diskussion von Gruppenchats bei Community-Manager*innen. Diese setzen die Regeln, die sich die Betreibenden von Social-Media-Seiten oder -Gruppen selbst geben, durch und achten auf die Einhaltung.

Der Bundesverband für Community-Management definiert Community-Management als die Tätigkeit, die aktive Follower*innenschaft eines Unternehmens oder einer Organisation in Austausch zu bringen. Community-Manager*innen moderieren dafür die Kommentarspalten von Social-Media-Kanälen und beantworten Mitteilungen und Anfragen der Nutzenden, die sie beispielsweise per Direktnachricht erhalten (Pein 2016, 4). Die Tätigkeit ist abzugrenzen von der Erstellung von Inhalten für Social-Media-Seiten (Content-Creation), von der plattformseitigen Content-Moderation und auch von einer dritten Form der Moderation: der ehrenamtlichen Moderation von hauptsächlich gemeinschaftsbasierten Strukturen wie Reddit-Foren, Gruppenchats auf Messenger-Plattformen und in Facebook-Gruppen.

Community-Management ist die Tätigkeit, die aktive Follower*innenschaft eines Unternehmens oder einer Organisation in Austausch zu bringen. Community-Manager*innen moderieren dafür die Kommentarspalten von Social-Media-Kanälen und beantworten Mitteilungen und Anfragen der Nutzenden.

Anders als die Themen Hate Speech (Geschke 2019; Wachs 2021; Hoven 2022), rechtsextreme und verschwörungsideologische Strategien zur Diskursmanipulation (Amadeu Antonio Stiftung 2017; Amadeu Antonio Stiftung 2020b; Fielitz 2020), Plattformpolitiken (Fielitz 2020) und diskursive Ansätze wie Gegenrede bzw. „Counter Speech“ (Jost et al. 2020; Hangartner et al. 2021) sind die Auswirkungen regulierender sowie diskursiver Moderation durch Community-Manager*innen noch nicht systematisch und multidisziplinär untersucht worden. Erste Untersuchungen liefern jedoch

Hinweise darauf, dass sich das Engagement lohnt und das Klima in Kommentarspalten durch moderierende Eingriffe besser wird (Kramp 2018; Ziegele 2021).

Diese Erfahrung machen auch wir in unserer täglichen Arbeit: Mit unseren Bildungsangeboten unter anderem zu den Themen Community-Management und Shitstorms, aber auch durch unser Engagement als Community-Manager*innen arbeiten wir im *Projekt Civic.net – Aktiv gegen Hass im Netz*¹ an der Schnittstelle von Demokratiebildung und Digitalisierung. Mit unseren Schulungen, Bildungsmaterialien und Publikationen (Amadeu Antonio Stiftung 2020a, 2022) bestärken wir zivilgesellschaftliche Organisationen in Berlin darin, sich proaktiv in digitalen Räumen zu äußern und dadurch die eigenen Anliegen wirkungsvoll einzubringen, sich gegen drohende Angriffe und Anfeindungen zu wappnen und Souveränität im Umgang mit hasserfüllten Kommentaren zu gewinnen.

Zwischen Regulation und Ermutigung: Was kann Seitenmoderation leisten?

Zu den Aufgaben und täglichen Herausforderungen vieler Community-Manager*innen zählt es, den konstruktiven und argumentbasierten Austausch zwischen engagierten User*innen – auch zu gesellschaftlichen Streitthemen – auf ihrer Seite zu fördern und zugleich diskriminierungssensible und wertebasierte Kommunikationsregeln aufzustellen, die konsequent durchgesetzt werden. Die meisten Betreibenden und auch Nutzenden von Social-Media-Seiten wünschen sich für die Kommentarspalten eine Atmosphäre, in der Austausch, Teilhabe, Verbindung und Ermächtigung möglich sind. Voraussetzung dafür sind insbesondere ein klarer Wertekompass seitens der Moderierenden sowie ausreichend Ressourcen für die Moderationsarbeit – dazu gehören neben Personal und Zeit auch Material wie Moderationsleitfäden als Entscheidungsgrundlage. Diese Voraussetzungen sind aber nicht in allen Organisationen gleichermaßen gegeben.

Zu den Maßnahmen, die Moderierenden zur Verfügung stehen, gehören einerseits regulierende Mittel. Unerwünschte Inhalte wie Hate Speech, Beleidigungen, Falschinformationen, Verschwörungserzählungen oder Spam haben einen einschüchternden Effekt auf andere User*innen und wirken destruktiv auf Diskussionen. Sie sollten daher begrenzt werden. Das einfachste und

Unerwünschte Inhalte wie Hate Speech, Beleidigungen, Falschinformationen, Verschwörungserzählungen oder Spam haben einen einschüchternden Effekt auf andere User*innen und wirken destruktiv auf Diskussionen. Sie sollten daher begrenzt werden.

¹ <http://www.civicnet.berlin>.

effektivste Mittel für Community-Manager*innen liegt im Löschen und Verbergen entsprechender Beiträge. Daneben sollten potenziell strafbare Inhalte von ihnen auch zur Anzeige gebracht werden – entweder durch eine Anzeige bei der Polizei oder durch eine Meldung bei zivilgesellschaftlichen Meldestellen wie *Respect!*, wo Jurist*innen die Strafbarkeit von Postings und Kommentaren einschätzen, die Beweissicherung übernehmen und sie schließlich auch zur Anzeige bringen.

Dass ein Post gelöscht oder verborgen wurde, sehen spätere Betrachter*innen jedoch nicht. Es ist deshalb ratsam, dass Moderierende regelmäßig in Kommentarform die Gründe für ihre Entscheidungen darlegen. So ein Post ist eine Warnung an andere potenzielle Hasskommentator*innen und zeigt der Community, dass eine Seite ihre Gesprächsregeln für einen offenen und wertschätzenden Diskurs durchsetzt. Sie senken gleichzeitig die Hemmschwelle für konstruktive Kommentator*innen und die sichtbare Präsenz der Moderation hat einen zivilisierenden Effekt auf die Kommentarspalten (Ziegele 2019).

Daneben kommt es aber immer wieder vor, dass Kommentare nicht eindeutig gegen Gesetze, den individuellen Wertekompass der*des Moderierenden, die öffentliche Netiquette oder interne Moderationsrichtlinien verstoßen – und trotzdem in Aspekten oder in Gänze problematisch sind. Das betrifft zum Beispiel solche Beiträge, die subtile menschenverachtende oder diskriminierende Aussagen enthalten, bei denen sich die Abwertung hinter rhetorischen Fragen oder persönlichem Erleben verbirgt, sowie Kommentare, die typische Störstrategien wie Whataboutism², Themenhopping³ oder Trolling⁴ enthalten. Dann können Mode-

Moderierende können mit Gegenrede Hassredner*innen in ihre Schranken weisen, indem sie beispielsweise auf die Normverletzung hinweisen, die Menschenfeindlichkeit benennen, Widersprüche aufdecken, Desinformationen erklären und auf die Werte der Seite oder Organisation verweisen.

² Whataboutism beschreibt das in Kommentarspalten häufig zu beobachtende Prinzip, Aussagen durch Gegenaussagen zu beantworten, die nichts mit dem eigentlichen Thema zu tun haben. Die Gegenaussage impliziert den Vorwurf, ein Thema würde absichtlich nicht besprochen. Deshalb wird diese Strategie als Ablenkungsversuch eingesetzt, um dem Gegenüber Doppelmoral vorzuwerfen.

³ Beim Themenhopping werden Streitthemen ohne Bezug zueinander aneinandergereiht. Das stiftet Verwirrung und macht es Moderierenden wie auch anderen Diskussionsteilnehmenden schwer, inhaltlich darauf zu reagieren.

⁴ Als Trolling wird das Prinzip bezeichnet, durch provozierende und unsachliche Kommentare emotionale Reaktionen hervorzurufen. Als klassisches Merkmal von Trollen gilt weiterhin, dass sie keine sichtbare eigene politische Agenda verfolgen. Das Ziel hinter dem Vorgehen: „Trolle“ wollen Kommentarspalten durch eine Vielzahl von Beiträgen dominieren, Moderierende und Nutzende zu Überreaktionen verleiten, einen sachlichen Austausch und so letztendlich einen Diskurs nach demokratischen Prinzipien verunmöglichen.

rierende mit Gegenrede („Counterspeech“) Hassredner*innen in ihre Schranken weisen, indem sie beispielsweise auf die Normverletzung hinweisen, die Menschenfeindlichkeit benennen, Widersprüche aufdecken, Desinformationen erklären und auf die Werte der Seite oder Organisation verweisen. Die Ziele einer solchen Gegenrede bestehen darin, die Betroffenen abwertender Aussagen in Schutz zu nehmen sowie mitlesenden Dritten Argumente zugänglich zu machen.

Für gelingende Diskussionen ist es neben den regulierenden Tätigkeiten aber auch wichtig, dass Moderierende die wünschenswerten Aspekte von Gesprächen innerhalb ihrer Kommentarspalten in den Blick nehmen und positive Kommunikation anregen, Nutzer*innen zum Austausch ermutigen und sachliche Diskussionen fördern. Dieser Aspekt der Seitenmoderation wurde gerade in problemfokussierten

öffentlichen und wissenschaftlichen Debatten lange Zeit nicht ausreichend gewürdigt. Inzwischen gibt es jedoch erste Forschungsergebnisse, die darauf hindeuten, dass eine Verlagerung des Fokus auf erwünschte, konstruktive oder bereichernde Beiträge zu besseren Online-Debatten führen kann.

Inzwischen gibt es erste Forschungsergebnisse, die darauf hindeuten, dass eine Verlagerung des Fokus auf erwünschte, konstruktive oder bereichernde Beiträge zu besseren Online-Debatten führen kann.

Die Kommunikationswissenschaftler*innen Marc Ziegele und Dominique Heinbach von der Universität Düsseldorf haben hierfür das Konzept der *Empowerment-Moderation* entwickelt, mit mehreren Moderationsteams großer und kleiner Medienhäuser erfolgreich in der Praxis getestet und die Ergebnisse im Rahmen ihres Whitepapers „Hallo liebe Community. Konstruktive Online-Debatten fördern durch bestärkende Moderation“ veröffentlicht (Ziegele und Heinbach 2021). Zunächst sei es zentral, dass Community-Manager*innen Präsenz und Interesse zeigen, Gespräche initiieren und die Nutzer*innen ernst nehmen. Moderierende sollten sich in erster Linie als Strukturgebende betrachten, die ähnlich einer Fernsehtalkrunde durch Fragen das Gespräch in Gang bringen, Orientierung bieten, zusätzliche Informationen liefern oder Ziele der Diskussion benennen. Moderierende sollten hierbei eine Vorbildfunktion einnehmen, indem sie anerkennend und respektvoll kommunizieren, statt herablassend, überheblich oder ironisierend.

Personen, die mitdiskutieren, orientieren sich häufig am Diskussionsstil und -inhalt der sichtbaren Kommentare. Deshalb hat es sich als besonders effektiv erwiesen, kurz nach Veröffentlichung eines Posts einen Eröffnungskommentar zu schreiben – damit Seitenmoderator*innen die Diskussionen und den Austausch zwischen Nutzenden anregen, aber auch, um den gewünschten Grundton für die Debatte zu setzen.

Facetten einer anspruchsvollen Tätigkeit

Um Seitenmoderation als zivilgesellschaftliche Intervention gegen Abwertung von Einzelnen und ganzen Gruppen sowie als aktive und bewusste Maßnahme zur Verbesserung von Online-Diskussion zu etablieren, ist aufseiten der Moderierenden einiges Wissen und Handwerkszeug vonnöten: Moderator*innen brauchen dafür aktuelle Kenntnisse über wichtige Akteure der verschiedenen

rechtsextremen und verschwörungsideologischen Szenen. Sie müssen wissen, wie menschenfeindliche Narrative strukturiert sind, mit welchen Techniken der Irreführung und Ablenkung Akteur*innen arbeiten und warum sie bei einigen Menschen so wirkungsvoll sind. Die Community-Manager*innen müssen geübt darin sein, abwertende Ideologien hinter vermeintlich harmlosen Aussagen oder rhetorischen Fragen zu entschlüsseln und sollen sich täglich gegen die dahinterstehenden Aussagen öffentlich im Namen ihrer Organisation oder Seite positionieren.

Community-Manager*innen müssen geübt darin sein, abwertende Ideologien hinter vermeintlich harmlosen Aussagen oder rhetorischen Fragen zu entschlüsseln.

Das zeigt, wie anspruchsvoll diese Tätigkeit ist – das komplette Wissensgebiet zu überblicken, fällt selbst ausgewiesenen Expert*innen nicht immer leicht. Gleichzeitig werden diese Aufgaben häufig von prekär Beschäftigten oder projektbasiert angestellten Personen „nebenbei“ erledigt. Viele bespielen im Rahmen ihrer Tätigkeit mehrere Social-Media-Präsenzen gleichzeitig, erstellen den Content und betreuen die Community. Das betrifft insbesondere lokale Initiativen, denen finanzielle sowie personelle Ressourcen fehlen, die jedoch gleichzeitig auf die Öffentlichkeit der Plattformen angewiesen sind, weil sie nur selten Erwähnung in der Presse finden und kostenpflichtige Anzeigen nicht infrage kommen. Aber auch große Medien, in denen täglich tausende Kommentare bearbeitet werden müssen, lassen die anspruchsvolle Arbeit häufig von Studierenden mit Teilzeitstellen erledigen.

Wie bei traditionellen Care-Berufen ist Community-Management zu erheblichen Teilen eine emotionale Tätigkeit. Es ist etwa sehr viel Fingerspitzengefühl nötig, um die Community in einer hitzigen Diskussion einzufangen, ohne sie durch Überregulierung oder eine moralisierende Haltung abzuschrecken. Es erfordert innere Stärke, auch bei anmaßenden oder unfreundlichen Kommentaren empathisch zu reagieren. Viel Selbstreflexion wird immer dann vorausgesetzt, wenn die Moderierenden eigene Fehler oder die der Redaktion eingestehen und entschuldigen müssen. Es geht beständig darum, Bedürfnisse von Nutzer*innen zu erspüren, Gefühle anderer aufzufangen und dafür zu sorgen, dass diese nicht in Beschimpfungen anderer Nutzer*innen münden, sondern diese in zivilere Bahnen zu lenken. Im Community-Management geht es also darum, dass sich Nutzende wohl und wertgeschätzt fühlen.

Die Aufzählung zeigt: Es ist schwer, diesem idealen Anforderungsprofil überhaupt zu entsprechen. Diese Tatsache, gepaart mit einem Bewusstsein für die Bedeutung ihrer Rolle, kann unter denen, die sie ausüben, zu Überforderung, Stress, einem Gefühl der Erschöpfung oder Angst vor dem Scheitern führen. Hinzu kommt: Die Arbeit ist auch deshalb enorm belastend, weil viele Community-Manager*innen konstant und für mehrere Stunden am Tag oder sogar mehrere Tage am Stück hassefüllten, zum Beispiel misogynen, rassistischen oder antisemitischen Aussagen ausgesetzt sind (Gräßer 2017, 60f.). Manche von ihnen sind zudem selbst von den abwertenden und gewaltförmigen Aussagen betroffen: Entweder, weil sie selbst der Gruppe angehören, die im Kommentar abgewertet wird. Oder weil die Kommentierenden ganz gezielt die Organisation, ihre Mitglieder und Community angreifen, beispielsweise wenn gezielt Mitarbeitende oder Engagierte einer Organisation rassistisch abgewertet werden, die als migrantische Selbstorganisation im Netz für die Rechte von Geflüchteten streiten.

Selbstfürsorge, eine Organisationskultur der gegenseitigen Unterstützung und angemessene psychologische Betreuung für diejenigen, die beruflich oder im Rahmen ihres Engagements mit Hass und Abwertungen umgehen müssen, sind maßgeblich für eine gesunde Bewältigung solcher Angriffe (Pein 2022; Leisenberg 2020).

Community-Manager*innen sind doppelt unsichtbar – in Kommentarspalten und gesellschaftlich

Gemessen an der hohen gesellschaftlichen Relevanz der Tätigkeit stehen für diese Aufgabe oft nur geringe Ressourcen zur Verfügung und das Tätigkeitsprofil ist verhältnismäßig unbekannt. Community-Management ist doppelt unsichtbar: Einerseits ist es in den Kommentarspalten unsichtbar, weil die meisten moderierenden Eingriffe darin bestehen dürften, dass Kommentare gelöscht oder verborgen werden – und nachfolgende Nutzer*innen nicht sehen können, dass ein Kommentar oder mehrere nicht mehr sichtbar sind. Nutzende dürften das nur in den seltensten Fällen hinterfragen, sie gehen wohl vielmehr davon aus, dass der verhältnismäßig harmonische oder gewaltfreie Zustand natürlich gegeben oder selbstverständlich ist.

Gemessen an der hohen gesellschaftlichen Relevanz der Tätigkeit stehen für diese Aufgabe oft nur geringe Ressourcen zur Verfügung und das Tätigkeitsprofil ist verhältnismäßig unbekannt.

Zweitens ist die Aufgabe auch auf gesellschaftlicher Ebene unsichtbar, weil sie nicht als eigenständige Arbeit anerkannt wird. Das ist beispielsweise innerhalb von Organisationen der Fall, wenn es zu wenig

Ressourcen in Form von Personal und Zeit gibt. Das trifft auf annähernd alle Organisationen zu, mit denen wir bisher im Rahmen unserer Arbeit sprachen. Die Betreuung der Kommentarspalten und Communitys wird weiterhin – anders als die Erstellung von „Content“ – gar nicht als anstehende Aufgabe anerkannt. Leitende Mitarbeitende gehen davon aus, das könne doch „nebenbei“ erledigt werden und berücksichtigen die Arbeit bei der Projektplanung nicht als zusätzliche Aufgabe. Dadurch bleiben Potenziale ungenutzt – das liegt auch daran, dass nicht alle Führungskräfte Social Media als relevanten Verbreitungsweg der eigenen Botschaften oder Produkte anerkennen. Oftmals mangelt es außerdem an Verständnis für grundlegende Funktionsweisen von Social Media generell sowie für Spezifika der einzelnen Plattformen. Dabei leisten Community-Manager*innen wichtige Beziehungsarbeit für ihre Organisationen mit der sympathisierenden und interessierten Follower*innenschaft.

Wie können Community-Manager*innen nun für eine größere Sichtbarkeit ihrer Moderationsleistung innerhalb der Kommentarspalten sorgen? – Indem sie sich nicht auf regulierende Praktiken beschränken, sondern einerseits, wie bereits oben beschrieben, ihre moderierenden Eingriffe transparent sichtbar machen. Und indem sie andererseits mit einem proaktiven Ansatz erwünschte Inhalte und konstruktive Beiträge durch eigene Kommentare anregen, hervorheben und fördern.

Auf Organisationsebene sind ausreichende Ressourcen eine Bedingung für gelingendes Community-Management. Dazu gehören neben Personal und Zeit auch Materialien und formalisierte Prozesse, die die – auch aus diskursethischer Sicht problematische – Individualisierung normativer Entscheidungen und eine Überforderung der Moderierenden verhindern. Dazu können insbesondere Moderationsleitfäden und Antwortbausteine beitragen, die als Entscheidungsgrundlage dienen. Das hat den Vorteil, dass mit solchen Hilfestellungen der Moderationsprozess formalisiert, Qualitätsstandards etabliert und damit der Druck für Moderierende, in strittigen Fällen immer neu entscheiden zu müssen, verringert wird.

Auch eine öffentlich zugängliche und gut sichtbar platzierte Netiquette sorgt für mehr Konsistenz, personenunabhängigere, transparent nachvollziehbare Moderationsentscheidungen und damit für mehr Sichtbarkeit für Community-Manager*innen. Weiterhin sollten bei Planungen zur Veröffentlichung von Inhalten auch die Moderationsleistungen mitbedacht werden – es macht beispielsweise wenig Sinn, besonders kontroverse Inhalte oder Themen kurz vor dem Wochenende zu posten, wenn dann keine angemessene Moderation der Kanäle

Besonders für kleine Organisationen und Redaktionen stellt die Sichtung und Regulierung der Kommentarspalten zu Nacht-, Wochenend- und Feiertagszeiten erfahrungsgemäß eine große Herausforderung dar.

sichergestellt werden kann. Bei solchen Themen sollten außerdem Anfeindungen und ein erhöhtes Auftreten von Hate Speech schon bei der Erstellung der Inhalte antizipiert, mögliche Antworten vorbereitet und die zusätzlich anfallende Arbeit und Belastungen eingeplant werden. Besonders für kleine Organisationen und Redaktionen stellt die Sichtung und Regulierung der Kommentarspalten zu Nacht-, Wochenend- und Feiertagszeiten erfahrungsgemäß eine große Herausforderung dar. Schlechtestenfalls leiden darunter wiederum die Community-Manager*innen, die aus einem Gefühl der Verantwortung für „ihre“ Seiten dann unbezahlte Mehrarbeit leisten, was nicht selten zu Erschöpfung von Engagierten und Angestellten führt. Organisationen können dem vorbeugen, indem sie auch diese Moderation außerhalb der üblichen Bürozeiten als Arbeitszeit einplanen – oder sollte das nicht möglich sein, die Kommentarspalten in den entsprechenden Zeiträumen schließen.

Auf politischer und gesellschaftlicher Ebene gibt es ebenfalls mehrere Ansatzpunkte, deren Zusammenwirken eine größere Sichtbarkeit und höheres Prestige für diese diskursprägende Tätigkeit versprechen. Presseberichterstattung kann dazu beitragen, den Beruf, seine gravierenden Herausforderungen und

Mehr wissenschaftliche Forschung sollte dazu beitragen, Wissen über die Wirkungen einzelner Maßnahmen auf die Diskussionsatmosphäre in Kommentarspalten, aber auch über die Bedeutung von Community-Management für den Diskurs zu aggregieren.

gesellschaftliche Bedeutung in der breiten Bevölkerung bekannter zu machen (Schroer 2023). Auf parlamentarischer Ebene können beispielsweise Kleine und Große Anfragen sowie Ausschusssitzungen die Arbeitsbedingungen oder Geschäftspraktiken der Digitalkonzerne problematisieren und zum Gegenstand politischer Auseinandersetzung machen. Interessenverbände wie der Bundesverband Community-Management arbeiten bereits an einer Etablierung, Differenzierung, Professionalisierung und Prägung der verschiedenen Social-Media-Berufsbilder in Deutschland. Darüber hinaus können sich Community-Manager*innen in selbstorganisierten Austauschrunden gegenseitig stärken, ermächtigen sowie über Herausforderungen und Probleme wie Stress sowie Gewalt- und Diskriminierungserfahrungen bei der Arbeit austauschen. Zivilgesellschaftliche Projekte wie *Civic.net*, die *Bildungsstätte Anne Frank* oder auch *empower communities* vernetzen Akteure, bilden sie fort und entwickeln gemeinsam mit ihnen solidarische Handlungspraktiken. Mehr wissenschaftliche Forschung sollte dazu beitragen, Wissen über die Wirkungen einzelner Maßnahmen auf die Diskussionsatmosphäre in Kommentarspalten, aber auch über die Bedeutung von Community-Management für den Diskurs zu aggregieren.

Ausblick: Künstliche Intelligenz im Kampf gegen Hate Speech?

Die Moderationsarbeit von Community-Manager*innen kann einen entscheidenden Beitrag dazu leisten, Online-Debatten zu verbessern und digitale Communitys zu schaffen, in denen die Mitglieder sich austauschen, ermächtigen und vernetzen. Neben einer großen Zahl individueller Kompetenzen und einem ausgeprägten Wertekompass benötigen sie dafür verbesserte Rahmenbedingungen in ihren Organisationen und ein höheres Maß an gesellschaftlicher Sichtbarkeit und Anerkennung. Social-Media-Plattformen setzen KI-gestützte Systeme längst im Alltag für die Einstufung jeglichen Contents ein (Gillespie 2020). Auch größere Onlineredaktionen, beispielsweise von Tageszeitungen, nutzen für die Moderation ihrer Social-Media-Kanäle Software, die KI-gestützt beispielsweise Spam und Hassrede erkennt. Die Tools können enorme Datenmengen verarbeiten und kategorisieren, um anschließend Entscheidungen zu unterstützen oder sogar automatisiert zu treffen und umzusetzen. Obwohl diese Technologie in jüngster Zeit rasante Fortschritte macht, kämpfen die Entwickler*innen Künstlicher Intelligenz nach wie vor damit, dass die Systeme Hate Speech nicht immer zuverlässig in all ihren Nuancierungen erkennen. Besonders Anspielungen, Ironie, subtilere Abwertungen oder spezifische soziale oder kulturelle Kontexte stellen die Künstlichen Intelligenzen weiterhin vor Herausforderungen (Ghafoor und Voß 2023). Doch selbst wenn eine menschliche Nachkorrektur weiter nötig bleiben sollte: Vielen überlasteten Community-Manager*innen ist damit geholfen, wenn die Technologie problematische Inhalte vorfiltert und sie dadurch Zeit gewinnen, um mehr Aufmerksamkeit auf die Förderung von zivilen und konstruktiven Diskussionen zu richten, bei denen die Nutzenden einen Mehrwert haben. Eine Aufwertung dieser meist unsichtbaren Arbeit geht damit jedoch nicht einher und muss in den Kommentarspalten, auf Organisationsebene, gesellschaftlich und politisch erkämpft werden.

Obwohl KI-gestützte Technologie in jüngster Zeit rasante Fortschritte macht, kämpfen die Entwickler*innen Künstlicher Intelligenz nach wie vor damit, dass die Systeme Hate Speech nicht immer zuverlässig in all ihren Nuancierungen erkennen.

Alina Darmstadt, 1. Staatsexamen, hat Bildende Kunst und Philosophie an der Goethe Universität Frankfurt am Main studiert und als Kunstvermittlerin gearbeitet. Seit 2017 arbeitet sie für die Amadeu Antonio Stiftung und leitet seit 2019 das Projekt Civic.net – Aktiv gegen Hass im Netz.

Antonia Graf, B. A., studierte Philosophie und Volkswirtschaftslehre an der Universität Mannheim und seit 2020 Politische Theorie im Master an der Goethe Universität Frankfurt am Main. Seit 2021 arbeitet sie als Social-Media- und Community-Managerin bei Civic.net – Aktiv gegen Hass im Netz.

Oliver Saal, M. A., studierte Geschichte an der Freien Universität Berlin. Für die Amadeu Antonio Stiftung arbeitet er seit 2017 – zuerst als Social-Media-Redakteur und inzwischen als Referent für digitale Themen bei Civic.net – Aktiv gegen Hass im Netz.

Teresa Sündermann, M. A., studierte Politikwissenschaften, Geschichte und Soziokulturelle Studien in Jena und Frankfurt (Oder). Sie arbeitet seit 2017 in der Amadeu Antonio Stiftung und ist seit 2020 Bildungsreferentin im Projekt Civic.net – Aktiv gegen Hass im Netz.

Literaturverzeichnis

- Amadeu Antonio Stiftung (2017a). Toxische Narrative. Monitoring rechts-alternativer Akteure. Berlin. Online verfügbar unter <https://www.amadeu-antonio-stiftung.de/wp-content/uploads/2018/08/monitoring-2017-1.pdf> (abgerufen am 25.10.2023).
- Amadeu Antonio Stiftung (2020a). Menschenwürde online verteidigen. 33 Social Media-Tipps für die Zivilgesellschaft. Berlin. Online abrufbar unter <http://www.menschenwuerde-online.de> (abgerufen am 16.10.2023).
- Amadeu Antonio Stiftung (2020b). Alternative Wirklichkeiten. Monitoring rechts-alternativer Medienstrategien: Berlin. Online verfügbar unter https://www.amadeu-antonio-stiftung.de/wp-content/uploads/2020/01/Monitoring_2020_web.pdf (abgerufen am 25.10.2023).
- Amadeu Antonio Stiftung (2022). Anti-Shitstorm-Kurs. Wie sie einen rechtsextremen Hassangriff gut überstehen: Berlin. Online abrufbar unter <https://www.amadeu-antonio-stiftung.de/publikationen/anti-shitstorm-kurs/> (abgerufen am 25.10.2023).
- Fielitz, Maik/Marcks, Holger (2020). Digitaler Faschismus. Die sozialen Medien als Motor des Rechtsextremismus. Berlin, Dudenverlag.
- Fielitz, Maik/Schwarz, Karolin (2020). Hate not Found. Das Deplatforming der extremen Rechten und seine Folgen. Jena. Online abrufbar unter https://www.idz-jena.de/fileadmin/user_upload/Hate_not_found/WEB_IDZ_FB_Hate_not_Found.pdf (abgerufen am 25.10.2023).
- Geschke, Daniel/Klaßen, Anja/Quent, Matthias/ Richter, Christoph (2019). #Hass im Netz: Der schleichende Angriff auf unsere Demokratie. Eine bundesweite repräsentative Untersuchung. Jena. Online abrufbar unter https://www.idz-jena.de/fileadmin/user_upload/Bericht_Hass_im_Netz.pdf (abgerufen am 25.10.2023).
- Ghafoor, Farangies/Voß, Oliver (2023). Deutschland-Chefs der Video-Plattform: „Nein, China hat keinen Einfluß auf Tiktok.“ Tim Klawns und Tobias Henning über ihre Strategie gegen Hassrede, Nachrichten von außerhalb der Filterblase und das Suchtrisiko für Jugendliche. Tagesspiegel vom 17.10.2023. Online zugänglich unter https://www.tagesspiegel.de/gesellschaft/medien_-ki/deutschland-chefs-der-videoplattform-nein-china-hat-keinen-einfluss-auf-tiktok-10630090.html (abgerufen am 25.10.2023).

- Gillespie, Tarleton (2018). *Custodians of the Internet. Platforms, content moderation and the hidden decisions that shape social media*. New Haven, Yale University Press.
- Gillespie, Tarleton (2020). Content moderation, AI, and the question of scale. *Big Data & Society* 7 (2).
- Gräßer, Lars (2017). Ein Interview mit Hildegard Stienen. Opferperspektiven, Strategien für die Verarbeitung von Hass im Netz und die Medien. In: Kaspar, Kai/Gräßer, Lars/Riffi, Aycha (Hg.). *Online Hate Speech. Perspektiven auf eine neue Form des Hasses (=Schriftenreihe zur digitalen Gesellschaft NRW)*, Marl, kopaed, 57–62.
- Hangartner, Dominik/Gennaro, Gloria/Alasiri, Sary (2021). Empathy-based counter speech can reduce racist hate speech in a social media field experiment. *PNAS* 118 (50).
- Jost, Pablo/Ziegele, Marc/Naab, Teresa K. (2020). Klicken oder tippen? Eine Analyse verschiedener Interventionsstrategien in unzivilen Online-Diskussionen auf Facebook. *Zeitschrift für Politikwissenschaft* (30), 193–217.
- Hoven, Elisa/Forschungsgruppe G/D/P (2022). *Hass im Netz. Ergebnisse einer Studie*. Leipzig. Online verfügbar unter https://www.gdp-group.com/fileadmin/user_upload/news/ms/hate_speech_22.pdf (abgerufen am 25.10.2023).
- Kloiber, Julia (2023a). „Content Moderatoren sind die Putzkräfte unserer Demokratie. Ohne ihre Arbeit würde es keine Social Media Plattformen geben. Thread in X.com (ehemals Twitter) am 19.06.2023. Online abrufbar unter https://twitter.com/j_kloiber/status/1670832273711562753, (abgerufen am 25.10.2023).
- Kloiber, Julia (2023b). *Social Media Content Moderators in Germany. Our Manifesto*. Online verfügbar unter <https://superr.net/2023/06/14/content-moderators-manifesto.html> (abgerufen am 25.10.2023).
- Kramp, Leif/Weichert, Stephan (2018). *Hass im Netz. Steuerungsstrategien für Redaktionen*. In: Landesanstalt für Medien NRW (Hg.). *Schriftenreihe Medienforschung*. Bd. 80. Leipzig, Vistas.
- Leisenberg, Sophia/Djassi, Barbara (2020). *Zeig mir den Schlüssel zur Bewältigung von Hass – Eine psychologische Perspektive. Ein Interview mit der Dipl.-Psych. Sophia Leisenberg*. Berlin. Online verfügbar unter <https://www.das-netz.de/zeig-mir-den-schlüssel-zur-bewaeltigung-von-hass-eine-psychologische-perspektive> (abgerufen am 25.10.2023).
- Newton, Casey (2019). *The Trauma Floor. The Secret Lives of Facebook moderators in America*. *The Verge* am 25.02.2019. Online verfügbar unter <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> (abgerufen am 25.10.2023).
- Pein, Vivian (2022). *Resilienz und Selbstfürsorge im Community Management*. Online verfügbar unter <https://der-socialmediamanager.de/selbstfursorge-achtsamkeit-community-management/> (abgerufen am 25.10.2023).
- Pein, Vivian (2016). *Bundesverband für Community Management e.V. Stellenprofil für Community Manager*. Online verfügbar unter <https://www.bvcm.org/wp-content/uploads/2021/11/160226-Whitepaper-Stellenprofil-Community-Manager.pdf> (abgerufen am 25.10.2023).
- Schroer, Fabian (2023). *Community Management. Für Respekt in der Kommentarspalte: taz vom 07.10.2023*, Online verfügbar unter <https://taz.de/Community-Management/!5962036/> (abgerufen am 16.10.2023).
- Wachs, Sebastian/Koch-Priewe, Barbara/Zick, Andreas (Hg.) (2021). *Hate Speech – Multidisziplinäre Analysen und Handlungsoptionen. Theoretische und empirische Annäherungen an ein interdisziplinäres Phänomen*. Wiesbaden, Springer.
- Welzel, Petra (2023). *4.000 Stunden Gewalt*. *ver.di publik* am 21.06.2023. Online verfügbar unter <https://publik.verdi.de/ausgabe-202304/4000-stunden-gewalt/> (abgerufen am 16.10.2023).
- Winkler, Gabriele (2015). *Care Revolution. Schritte in eine solidarische Gesellschaft*, Bielefeld, transcript.
- Ziegele, Marc/Jost, Pablo/Friefß, Dennis/Naab, Teresa (2019). *Aufräumen im Trollhaus. Zum Einfluss von Community-Managern und Aktionsgruppen in Kommentarspalten*. DIID-Précis.
- Ziegele, Marc/Heinbach, Dominique (2021). *Hallo liebe Community! Konstruktive Online-Debatten fördern durch bestärkende Moderation*. Düsseldorf. Online verfügbar unter https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Foerderung/Forschung/Dateien_Forschung/WhitePaper_HalloLiebeCommunity_DIGITAL.PDF (abgerufen am 16.10.2023).