

Online-Diskussionskultur - Stand der Technik

Lucas Braeschke

TU Dresden

Email: l.braeschke@outlook.com

Zusammenfassung—Online wird die Diskussion überschattet von „Hatespeech“, „Toxic Communitys“ und anderen Phänomenen. Dieser Aufsatz stellt einige Probleme des Diskurses im Web vor und beschreibt ihre Besonderheiten. Zunächst wird ein kurzer Blick auf die Entwicklung dieser in den Anfängen des Internets geworfen. Danach werden verschiedene Lösungsstrategien vorgestellt, welche schon im Einsatz oder aber in der Entwicklung sind. Diese sollen dabei helfen den Dialog in den verschiedensten Foren und Kommentarbereichen zu verbessern. Dazu werden diese Ansätze näher vorgestellt und nach ihren Einsatzmöglichkeiten bewertet. Zum Schluss erfolgt ein Vergleich der unterschiedlichen Maßnahmen, um mögliche Methoden aufzuzeigen, welche eingesetzt werden können, um die Online-Diskussionskultur wieder zu verbessern.

I. EINLEITUNG

Das Miteinander und die Diskussion im Internet sind geprägt von „Hatespeech“, „Shitstorms“ und „Toxic Communitys“. Dies hat zur Folge, dass Dispute, Kommentarbereiche und Diskussionen online einen sehr schlechten Ruf haben. Mittlerweile sind antisemitische, rassistische und andere abwertende Kommentare auch auf großen Plattformen wie etwa YouTube zum Alltag geworden.

Bemerkenswert dabei ist das Ausbleiben fast jeglicher Konsequenzen auf ein solches Handeln. Durch die Anonymisierung ist vielen Nutzern die Person hinter dem virtuellen Gegenüber nicht mehr deutlich. Dies enthemmt die Aussagen, welche in einem persönlich Gespräch nur selten geäußert würden. Hinzu kommen die grenzenlosen Möglichkeiten, welche sich im World Wide Web bieten.

Dabei wurde noch nie so viel über den Einzelnen preisgegeben wie im Netz. Konträr dazu steht jedoch auch, dass es keinen anderen Raum gibt, in welchem so anonym agiert werden kann wie im Internet. Aus diesem Paradox erwächst eine völlig neue Möglichkeiten der Selbstdarstellung. Es ist für viele enorm wichtig geworden, wie sie sich auf Social Media etc. darstellen. Durch die Anonymisierung können nur die Aspekte präsentiert werden, welche in die eigene Selbstdarstellung passen und somit das virtuelle Ich nach eigenen Wünschen formen.

Für die Diskussionen hat dies zur Folge, dass diese mit einer unvergleichlichen Härte und Unnachgiebigkeit geführt wird. In Folge dessen geht es meist mehr darum, den eigenen Standpunkt durchzusetzen als einen offenen und sachlichen Diskurs zu führen. Hinzu kommt, dass Nutzer existieren, welche nur aus der Freude möglichst viel Unruhe zu stiften, absichtlich versuchen die Diskussion zu sabotieren.

Ein weiteres Problem ist die schriftliche Form. Besonders Ironie oder Sarkasmus sind teilweise schwer zu erkennen.

Durch die Interpretationsmöglichkeiten entstehen oft Konflikte, welche anhand solcher Missverständnisse entflammen können.

Im Folgenden soll näher auf die Geschichte der Diskussionskultur im Internet und die verschiedenen Probleme in der Diskussion dargestellt werden. Anschließend werden verschiedene Ansätze vorgestellt, welche sich noch in der Entwicklung oder aber auch schon im Einsatz befinden, um ein geregeltes Miteinander im Web zu ermöglichen.

II. EINE GESCHICHTE DES HASSES

In der heutigen Darstellung erscheint es, als ob sich Begriffe wie „Hatespeech“ in den Diskussionen des World Wide Webs gebildet haben. Betrachtet man jedoch die Geschichte von Hass und Volksverhetzung, so wird klar, dass die Ursprünge tief in der Geschichte des Menschen verankert sind. Ziel ist und war es schon immer, ein anderes Individuum, eine Gruppe oder Minderheit über eine Form von Kommunikation zu unterdrücken [1, Seiten 48-62]. Dafür lassen sich in der Menschheitsgeschichte zahllose Beispiele finden, wie etwa die Bezeichnung afrikanischer Sklaven als „Boy“, die antisemitischen Darstellungen von Juden im dritten Reich oder die Darstellung des Islams als radikale Terroristen in manchen heutigen Medien (auch außerhalb des Internets). Aus diesem Grund ist es wenig verwunderlich, dass dieses tief verwurzelte Übel auch im Internet Einzug gehalten hat. Was jedoch verwundert, sind die Häufigkeit und der Grad an Hass in den Beleidigungen, welcher zumindest in Debatten und anderer öffentlicher Kommunikation in der heutigen Zeit nicht wiederzufinden ist.

Ein entscheidender Faktor dabei ist die Anonymisierung der Nutzer. Durch die Beschränkungen der Kommunikation und die damit einhergehenden, fehlenden Sinneseindrücken von anderen Personen, sorgt für eine veränderte Wahrnehmung anderer Nutzer. Dies kann „prosoziales aber auch antisoziales Verhalten“ [2, s.371 Tab. 13.5] fördern. Schon in den Anfängen des Internets, in den ersten Netzwerkverbänden, wurden andere Teilnehmer attackiert und es gab erste Auffälligkeiten wie etwa „guide to flaming on BBS.“ [3, Seite 2]. Mit der rasanten Ausbreitung des Internets und der damit einhergehenden leichteren Zugänglichkeit, vergrößerten sich auch diese Gruppen und bildeten eigene Sub-Kulturen. Mit dem Voranschreiten zum Web 2.0 entstehen aus diesen Gruppen verschiedenste Communitys, welche die neuen Interaktionsmöglichkeiten für ihre Zwecke nutzen. Ein prominentes Beispiel ist etwa „4chan“, welches 2003 startete und berühmt berüchtigt für verschiedene Trolls und Memes ist.

Das Problem, welches sich jedoch aus diesen neuen Möglichkeiten ergibt, ist dass sich Hasskommentare und „Trolls“ nicht mehr auf einen Bereich begrenzen. Sondern überall wo es nicht durch scharfe Regeln und Aufsicht unterbunden wird, fast ungehindert ausgeübt werden können[4]. Dabei haben besonders größere Plattformen wie etwa YouTube, Facebook oder Twitter teils immense Probleme[5] dem Verhalten ihrer Nutzer Herr zu werden.

III. AUSPRÄGUNGEN

Besonders durch das Smartphone und dem damit jederzeit möglichen Zugang ins Internet, hat sich unser soziales und somit auch politisches Leben in den letzten Jahren enorm gewandelt. Durch die immer stärkere Vernetzung und dem sich Auflösen der Grenzen zwischen realer und virtueller Welt, bekommt die Selbstdarstellung in dieser einen immer stärkeren Stellenwert. Anders herum beeinflussen soziale Medien und Diskussionen uns stärker, als kaum ein anderes Medium, weshalb es kaum verwunderlich ist, dass die Auswirkungen auf unser alltägliches Leben immens sind. Dabei zeigt sich aber, dass die Interaktion über das Web an vielen Stellen durch Hass und Wut geprägt wird. Es gibt viele Problemstellen und Phänomene, welche genutzt werden um die eigene Meinung möglichst aufmerksamkeitswirksam durchzusetzen.

A. Shitstorm

Der Shitstorm, auch bekannt als Flameware, ist der Alptraum eines jeden Marketing- oder Pressesprechers. Nach dem Duden ist dies ein „Sturm der Entrüstung in einem Kommunikationsmedium des Internets, der zum Teil mit beleidigenden Äußerungen einhergeht“[6]. Ein Begriff welcher 2011 zum Anglizismus des Jahres gewählt wurde.

Das Besondere dabei ist jedoch die scheinbar neu entdeckte Macht der Konsumenten, welche in der Masse selbst große Hersteller in die Schranken weisen können. Jene bedienen sich meist Hatespeech, um ihrer Verärgerung Ausdruck zu verleihen. Selbst geringfügige Anlässe können die Wut der Verbraucher auslösen, wie etwa der Schokoladen-Weihnachtsmann, welcher bei Penny nun „Zipfelmann“[7] heißt.

Dabei muss jedoch gesagt werden, dass die Auswirkungen oft eher gering sind. So ist meist der mediale Schaden höher als der Finanzielle, das liegt unter anderem daran, dass sich viele Nutzer gern aufregen; ihnen es dann aber doch schwer fällt, ihre Vorsätze auch in der realen Welt in die Tat umzusetzen. Im Fall von Penny hatte der Shitstorm sogar teilweise eine verkaufsanregende Wirkung und so dem Unternehmen durch die kostenlose Werbung Gewinn gebracht.

B. Fake News

Der Begriff Fake News bezeichnet „Falschmeldungen oder irrtümlich als Fakt dargestellte Neuigkeiten“[8]. Dieses Phänomen wurde in den vergangenen Jahren immer aktueller und besonders im US-Wahlkampf stark thematisiert. Diese Nachrichten werden insbesondere in Social Media verbreitet und stellen Plattform-Betreiber, aber auch Nutzer vor große Probleme.

Für die Betreiber ist es schwierig die Verbreiter solcher Nachrichten dauerhaft aus ihren Systemen zu vertreiben und für die Nutzer ist es schwierig die Informationen zu validieren. „In der Endphase des US-Wahlkampfes wurden die 20 erfolgreichsten Falschmeldungen öfter geteilt, gelikt und kommentiert als die 20 erfolgreichsten Berichte seriöser Medien.“[8]. Dies zeigt auf, welchen Stellenwert solche Falschmeldungen haben können.

Dabei können die Fake News nach [9], in drei Kategorien eingeteilt werden „Hoaxings“, „Ernsthafte“ und „Satirische Falschmeldungen“[9].

Hoaxings: sind Falschmeldungen welche auch als „Urban Legend“ bezeichnet werden. Dabei handelt es sich meist um eine Nachricht oder ein Gerücht, welche irrtümlicherweise als valide angesehen werden. Diese werden dann von Nachrichtenjournalen und vor allem in Social Media verbreitet und meist wenig hinterfragt. Ein Beispiel, in welchem viele solcher Gerüchte auftraten, ist die Flüchtlingskrise in Deutschland. In dieser kam es vermehrt zu Falschmeldungen, vor allem in Social Media, über angebliche Gewaltverbrechen von Flüchtlingen[10].



Abbildung 1. hoaxmap.org sammelt bestätigte Hoaxings und stellt diese nach Regionen dar.

Ernsthafte Falschmeldungen: sind all jene Meldungen, welche bewusst überspitzt falsche Zusammenhänge darstellen oder einfach nur die Fakten nicht korrekt wiedergeben. Dabei kann das Motiv ganz unterschiedliche Formen annehmen. Entweder für die erhöhte Aufmerksamkeit und somit Reichweite, was z.B. der Bild-Zeitung vorgeworfen wird (wie etwa in diesem Artikel [11]), oder aber auch, um gezielt Desinformationen und Hass zu schüren. Letzteres kann im Web mit in die Kategorie Troll gezählt werden.

Satirische Falschmeldungen: zeichnen sich durch die Prämisse und den komödiantischen bzw. satirischen Inhalt aus. Der Sinn und Zweck ist hier nicht, absichtlich Falschmeldungen zu verbreiten, sondern das Format der Neuigkeiten ins Absurde zu treiben, um mit dieser den Leser zu unterhalten. Der Inhalt ist dabei oft überspitzt dargestellt und es fällt dem Leser meist nicht weiter schwer die Meldung als Satire zu erkennen. Ein gutes Beispiel wären die Artikel des Postillon, welche alle in dieses Muster fallen.

Für den Nutzer ist es oft schwer, zwischen wahrer Nachricht und Fake News zu unterscheiden. Dabei ist dies ein Problem

der Presse allgemein. Denn diese bürgen mit ihrem Namen und ihrer Authentizität für die Korrektheit der Meldung. Zu Zeiten von Suchalgorithmus Optimierung wird es jedoch wichtiger, möglichst viele News und als erster zu publizieren. Hinzu kommt, dass oft Nachrichten anderer aufgegriffen und verbreitet werden, ohne noch einmal die Quellen gründlich zu prüfen. Das wiederum verunsichert den Leser und macht es für ihn sehr schwer den Wahrheitsgehalt einer Meldung zu überprüfen. Was wiederum sein Vertrauen –besonders wenn dieses enttäuscht wird– verringert.

C. Filterblase

Die Filterblase (engl. filter bubble) oder auch Echokammer, beschreibt das ausschließliche konsumieren von Informationen und Meinungen, welche konform mit dem eigenen Weltbild stehen. Besonders im Internet lässt sich dieser Effekt deutlich erkennen.

Social Media und Suchergebnisse werden durch Algorithmen angepasst und individuell vorbereitet (vgl. [12] und [13]). Dies hat zur Folge, dass die Realität teilweise verfälscht wird. Denn durch die einseitige Darstellung und die ständige Bestätigung der eigenen Meinung, geht die diverse Darstellung mit verschiedenen Blickwinkeln verloren. Durch die personalisierte Vorauswahl der Algorithmen, gelangen andere Meinungen und Darstellungen einfach nicht mehr in den eigenen Kosmos. Es benötigt also eine aktive Handlung, sich von seinem eigenen Weltbild zu entfernen und sich anderweitig zu informieren.

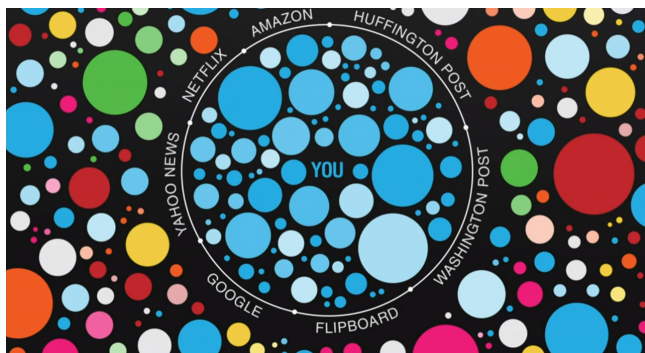


Abbildung 2. Die Filterblase (Quelle: cdn-images-1.medium.com)

In gewisser Weise ist die Echokammer kein neuer Effekt. Schon immer hat die Wahl von Freunden und der Konsum von bestimmten Medien die Bedingungen der Echokammer erfüllt. Das Internet und die neuen Möglichkeiten der Kommunikation und der Verbreitung von Informationen hat diesen Effekt jedoch noch einmal verstärkt.

Dabei steht dies im besonderen Zusammenhang mit den Fake News. Denn innerhalb von Social Media etc. entstehen und verbreiten sich besonders Hoaxes sehr schnell. Das Problem dabei ist, dass diese meist nicht als solche demaskiert werden. Durch die Selektierung der Algorithmen und dem Effekt der Filterblase werden so andere Darstellungen und Blickwinkel oft nicht angezeigt.

D. Social Bots

Ein Social Bot ist nach [14] ein Account in den Sozialen Netzwerken, welcher nicht von einem Menschen, sondern von einem Bot verwaltet wird. Diese Bots werden genutzt, um für bestimmte Beiträge, Meinungen oder Artikel, Aufmerksamkeit und scheinbares Interesse zu generieren. Durch die Algorithmen, mit welchen die Sozialen Netzwerke arbeiten, wird dann die Nachricht höher priorisiert und somit die Sichtbarkeit erhöht, was wiederum noch mehr Aufmerksamkeit erzeugt. Alle großen Anbieter verbieten Bots, jedoch ist es fast unmöglich, solche Praktiken aus den Netzwerken auszusperren.

Genutzt werden die Social Bots, um Nachrichten oder Statements, meist mit einem politischen Hintergrund, schnell zu verbreiten und eine möglichst große Zielgruppe zu erreichen. Im US-Wahlkampf sind diese sowohl auf Seiten von Donald Trump, als auch von Hillary Clinton zum Einsatz gekommen.

Das Problem ist, dass die eigentliche Nachricht dabei zur Nebensache verkommt. Es wird immer wichtiger Beiträge und Meldungen auf Reichweite zu optimieren, um möglichst hoch in den Suchmaschinenanfragen priorisiert zu werden, anstelle mit Inhalten zu überzeugen. Dabei stehen die Bots noch am Anfang ihrer Möglichkeiten. Momentan ist es noch relativ leicht einen solchen zu entlarven, jedoch könnte sich dies in den nächsten Jahren ändern. Dabei könnte sich zum Einen ändern, wie sie Inhalte verbreiten, als auch in welchen Bereichen sie eingesetzt werden können. Das wird sich mit voranschreitender Entwicklung und immer leistungsstärkeren Algorithmen stark verändern und das Maß in welchem sie eingesetzt werden, noch erhöhen.

E. Toxic Communitys

Der Begriff Toxic hat sich aus dem Englischen etabliert und wird mittlerweile oft im Zusammenhang mit Kommentaren gebraucht. Dieser vereint sowohl die Hasskommentare aber auch Hetze, Trolls und einfach negative Kommentare als Oberbegriffe.

„Toxic Communitys“ ist in diesem Zusammenhang ebenfalls verstärkt aufgetaucht. Dieser wird vor allem im Gaming, aber auch außerhalb gebraucht, um eine Community, wie etwa eine Spielerschaft eines MOBAS oder auch eine Nutzergruppe auf Plattformen wie YouTube oder Facebook zu bezeichnen. Gemeint ist damit, dass ein Großteil der Nutzer sehr schnell mit negativen Kommentaren und Nachrichten auf Außenstehende, Neulinge, Vertreter unpopulärer Meinung und sogar den Entwickler oder Plattformbetreiber reagiert.

Solche Portale gelten als Problemfälle des Internets, in welchen ein höfliches Miteinander und eine offene Diskussion nicht mehr möglich sind. Das Problem besteht darin, dass es meist nicht mehr möglich ist in einer Community, welche einmal einen solchen Ruf erlangt hat, die Diskussionskultur wieder herzustellen. Das liegt zumeist daran, dass die gemäßigteren Nutzer, welchen ein solches Klima nicht gefällt, abgewandert sind oder sich nicht mehr beteiligen. Bei den

anderen Nutzern entsteht dann der Eindruck, dass Hatespeech in der Community zum Ton gehört und sich die Nutzer untereinander –ähnlich wie in einer Filterblase– gegenseitig darin bestärken.

Als Gegenmaßnahme bleibt dann oft nur noch hartes Durchgreifen, welches jedoch oft auf Unverständnis trifft. Dies kann sogar den gegenteiligen Effekt erzeugen, sodass sich die Community gegen die Plattformbetreiber richtet. Die Nutzer fühlen sich in einem solchen Fall oft in ihrer freien Meinungsäußerung eingeschränkt und werfen vor allem Moderatoren und Betreibern „Zensur“ vor.

IV. AUSWIRKUNGEN

Beobachtet werden kann, dass die freie Diskussion überschattet wird und ein offener Diskurs nicht mehr uneingeschränkt möglich zu sein scheint. Leidtragende sind jedoch meist keine Unternehmen oder große Persönlichkeiten des öffentlichen Lebens, wie zunächst vermutet werden kann. Für diese agieren meist Marketing Agenturen oder beraten diese. Außerdem kann festgestellt werden, dass sich die Auswirkungen für diese als eher gering einschätzen lassen können.

Ein Beispiel dafür ist etwa Call of Duty: Infinite Warfare, welches nach der Ankündigung eine riesige Kritik einfahren musste und dessen Trailer auf YouTube als eines der schlecht bewerteten Videos aller Zeiten gilt. Dennoch hat sich der Publisher Activision nicht genötigt gefühlt, etwas an seinem Konzept zu ändern [15]. Die Verkaufszahlen jedoch geben Activision Recht [16] und Call of Duty spielte trotz des enormen Hasses im Vorfeld wieder riesige Gewinne ein.

Die eigentlichen Verlierer stehen eine Stufe tiefer. Keine großen Politiker oder Firmen, sondern leitende Entwickler oder YouTuber. Diese bekommen den Hass wesentlich unmittelbarer und persönlicher mit. Sie sind namentlich bekannt und stellen in gewisser Weise eine Identifikationsfigur dar, welche für Handeln steht. Auf diese kann der Hass wesentlich zielgerichteter wirken, als zum Beispiel auf ein so gewaltiges Konstrukt wie ein Entwicklerstudio. Dies hat zur Folge, dass vor allem die Kreativen bei heiklen Themen mit einer gewaltigen Masse an Anfeindungen zurechtkommen müssen.



Abbildung 3. Offener Brief von Charles Randall auf Twitter.

Ein Beispiel wäre Anita Sarkeesian, eine feministische Videospielaktivistin, welche auf YouTube eine Videoreihe „Tropes vs Women“ veröffentlicht. Das Missfallen einiger Nutzer entlud sich in einer Welle von sexistischen Hasskommentaren

und sogar Morddrohungen[17]. Dabei ist dieser Shitstorm vor allem durch gezielte persönliche Anfeindung gekennzeichnet. Auch andere Kreative leiden darunter. Der Spieleentwickler Charles Randall beschreibt in einem offenen Brief auf Twitter [18] die Situation aus seinem Blickwinkel. Er berichtet von der Angst –seiner eigenen und der anderer Entwickler–, offen über ihre aktuellen Projekte zu sprechen, aus Furcht vor Repressalien Seitens der Community. Dabei bestehe laut Randall aus Entwicklersicht der Wunsch und auch der Bedarf nach einer offenen Kommunikation. Doch die Gefahr, mit einem Feature oder nur einer Aussage, Ziel des Missfallens der Fans zu werden, sei zu groß.

Dadurch schaden die Nutzer sich eigentlich selber, indem sie so aggressiv im Internet vorgehen. Insgesamt kann festgestellt werden, dass vor allem jene Menschen besonderes unter der Diskussionskultur im Internet leiden, welche kreative Arbeit leisten. Dabei werden besonders die zur Zielscheibe, die evtl. neue und kontroverse Ideen umsetzen und meist in einer neuen Form Inhalte erschaffen.

V. GEGENMASSNAHMEN

Daraus ergibt sich die Frage, was überhaupt unternommen werden kann, um dieser Entwicklung von Shitstorms und Toxic Communitys entgegenzuwirken. Dafür gibt es verschiedenen Ansätze, welche im Folgende genauer vorgestellt werden sollen.

A. Counterspeech

Die offensichtlichste Gegenmaßnahme gegen Hasskommentare und ähnliches, ist es einfach einer Meinung zu widersprechen. Dabei kann dies durch eine Gegenargumentation und Gegen-Posting erfolgen. Etwa mit einem Verweis auf eine externe Quelle, welche Fake News entlarvt und damit versucht wird den Kommentarschreiber aus seiner Filterblase zu reißen. Auf dieses Konzept berief sich u.a. Facebook. „Facebook möchte das Konzept Counterspeech in Deutschland etablieren und aktive Gegenrede weiter fördern. Counterspeech setzt auf den Widerstand der Gesellschaft und jedes Einzelnen im Kampf gegen Hassrede“(zitiert nach [19]).

Eine eco Umfrage „zeigt jedoch, dass sich lediglich 14 Prozent der Teilnehmer mit einem eigenen Kommentar, auf rassistische Posts im Internet reagieren würden. Beinahe die Hälfte (45 %) der Befragten gaben hingegen an, rassistische Hassbotschaften im Internet komplett zu ignorieren“[20].

Widerspruch zu äußern, kann jedoch auch eine Kontroverse anfeuern. „Es kann sowohl für die Legitimität der Inhalte als auch für die Lebensdauer der Kontroverse Folgen haben“ [1]. Denn Counterspeech kann zur Folge haben, dass dadurch noch mehr Ereignisse generiert werden. Aus dem erzeugten Konflikt erwächst so noch mehr Konflikt, welcher wiederum Folgeereignisse generiert (vgl. mit [1]).

Andererseits ist es ebenfalls wichtig, die Opfer von Anfeindungen nicht alleine zu lassen und mit einem Kommentar und dem damit symbolischen Einschreiten kann eine Solidarisierung erzeugt werden. Der Widerspruch zeigt



Abbildung 4. Ausschluss eines Nutzers von Facebook.

den Verbreitern von Hass, dass ihre Meinung nicht akzeptiert wird. Beim Verfasser könnte sonst der Eindruck entstehen, dass seine Meinung stillschweigend akzeptiert wird, wenn es kein Contra auf Hasskommentare gibt. Dies kann dabei helfen die Filterblase zu durchbrechen und ist umso effektiver, je näher der Kontakt zur entsprechenden Person ist. Je anonym der Kontakt jedoch ist, umso geringer ist die Möglichkeit Einfluss zu nehmen(vgl.[19]).

Counterspeech ist kein Allheilmittel, der Wirkungsbereich beschränkt sich auf enge Freunde. Die anonymen Nutzer im Internet lassen sich durch ein Contra kaum beeindrucken. Im Zweifelsfall kann ein Einspruch noch einmal einen Konflikt anfeuern. Der eigentliche Nutzen ist Schadensbegrenzung. Es hilft Betroffenen zu zeigen, dass diese nicht alleine sind und Grenzen aufzuzeigen. Dass jedoch die Nutzer, wie etwa Facebook es sich wünscht, damit Unerwünschtem entgegenzutreten und somit auch die Diskussionskultur verbessern, ist kaum praktikabel.

B. Moderatoren

Moderatoren kümmern sich in Foren, Sozialen Netzwerken etc. darum, dass dort eine offene Diskussion möglich ist. Unternehmen stellen dafür oft Community Manager ein, welche für sie diese Arbeit professionell übernehmen. Hauptaugenmerk liegt dabei darauf, die Kommunikation mit Nutzern aufrecht zu erhalten und diesen so eine Möglichkeit zu geben, ihre Wünsche zu äußern.

Moderatoren sind auch dafür verantwortlich, dass sich an eine vorgegebene Diskussionskultur gehalten wird. Dafür folgen sie den aktuellen Diskussionen und versuchen diese teilweise auch in eine gewisse Bahn zu lenken. Außerdem besteht oft eine Meldfunktion, um einen Moderator auf einen Beitrag aufmerksam zu machen. Wenn ein Nutzer sich nicht an die Regeln hält, kann durch unterschiedliche Möglichkeiten eingegriffen werden.

Es besteht die Möglichkeit einfach nur den Post zu löschen, oder aber den Nutzer als Strafe für eine Weile zum Schweigen zu bringen, so dass dieser sich nicht mehr an der Diskussion beteiligen kann. Oder aber der entsprechende Nutzer kann für eine gewisse Zeit oder ganz verbannt werden.

Dabei muss der Moderator Fingerspitzengefühl beweisen und unterscheiden und erkennen, wann jemand nur ein wenig

über die Stränge geschlagen hat oder aber versucht nur Ärger zu machen. Auch das komplette Thema zu schließen, wenn dieses aus dem Ruder läuft, kann eine Maßnahme sein. In jedem Fall ist es wichtig, dass schnell reagiert wird und der Community auch mitgeteilt wird, weshalb die jeweilige Aktion durchgeführt wurde, um weitere Aggressionen auf diese zu verhindern [21].

Es besteht die Frage, wie wirksam solche Maßnahmen sind. Es wird damit gerechnet, dass dem Nutzer sein Profil so wichtig ist, dass er alles versucht, um dieses zu erhalten. Es ist jedoch gängige Praxis, dass Verbreiter von Hass und Trolls, sich neue Profile nur zu diesem Zweck erstellen, sodass ein Ban sie nicht weiter kümmert, da sie jederzeit ein neues Profil erstellen können. Außerdem empfinden viele Nutzer das Löschen von Beiträgen etc. als Zensur und dies kann z.B. im Falle eines Shitstorms die Gemüter noch weiter erhitzen.

Der Aufwand, welcher durch die Moderation erfolgt ist enorm. Die New York Times gab bekannt, dass 14 Moderatoren ungefähr 11,000 Kommentare am Tag begutachten. Damit eine Moderation für jeden Artikel gewährt werden kann, können Kommentare trotzdem nur unter 10% der Artikel aktiviert werden. (siehe [22]). Dieser Aufwand könnte sogar noch größer werden. 2017 gibt es in Deutschland einen Gesetzesentwurf, welcher Soziale Netzwerke dazu verpflichtet, eindeutig gesetzwidrige Beiträge innerhalb von 24 Stunden zu löschen, da ansonsten hohe Strafen drohen.

Die Moderatoren können nicht verhindern, dass eine Diskussion außer Kontrolle läuft. Sie können nur versuchen, diese einzudämmen. Im Prinzip ist es ihnen nur möglich, zu erahnen wo der nächste Brandherd entsteht und dann versuchen diesen zu löschen. Sie können kurzfristig etwas bewirken; um jedoch langfristig etwas zu verbessern, ist der Aufwand, welcher von Moderatoren betrieben werden müsste, gigantisch.

C. Perspektive

Perspektive[23] ist ein neuer Service von Google. Dieser bietet eine API, welche Forenbetreiber und Andere dabei unterstützen soll „toxic comments“ zu filtern, bevor diese abgegeben werden können. Dazu nutzt Perspektive das gewaltige Potenzial des Maschinenlernens, um damit einer K.I. das Erkennen von Hasskommentaren etc. beizubringen. Genutzt wird dazu die hauseigene Open-Source Bibliothek „Tensorflow“.

Auf der Projektseite kann Perspektive getestet werden. Dabei gibt es die Möglichkeit, einen Kommentar abzugeben und als Feedback wird ein Prozentsatz angegeben, welcher die Wahrscheinlichkeit anzeigt, mit der der Kommentar „toxic“ ist. Über die API kann mit Hilfe dieses Prozentwertes eingestellt werden, ab wann Kommentare nicht mehr angezeigt werden. Im Prinzip ein Löschen des Kommentars im Voraus.

Zum Einsatz kommt Perspektive unter anderem in der New York Times. Dort wird es testweise für die Auswertung der

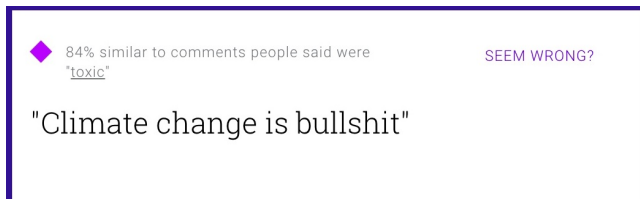


Abbildung 5. GUI von Perspektive.

Kommentare genutzt. Ziel ist es, vor allem die Modelle zu verbessern. Dafür gibt es eine Feedbackfunktion, mit welcher ein Nutzer, der einen Kommentar als „toxic“ empfindet, diesen an Perspective übermitteln kann. Der Betreiber kann dabei auswählen, ob die Kommentare dabei dauerhaft gespeichert werden dürfen. Diese Funktion liefert der K.I. Lerndaten, an Hand welcher dieser dann zukünftige Kommentare schätzt.

Wikipedia hat sich ebenfalls an dem Projekt beteiligt. Insgesamt wurden 160000 Anmerkungen von 5000 Wikipedia Autoren eingeschätzt. Dabei wurde jeder Kommentar von 10 Autoren bewertet. Momentan wird von Perspektive nur Englisch unterstützt, weitere Sprachen sollen aber mit fortlaufender Entwicklung noch integriert werden.

Es gibt jedoch bereits Kritik am Projekt [24]. Das Problem besteht darin, dass die K.I. nicht kontextsensitiv genug ist, um die Kommentare richtig zu bewerten. Jedoch gibt es Fälle, in welchen es bestimmte Wörter gibt, die in einem bestimmten Kontext beleidigend sind, in einem anderen jedoch z.B. nur die persönliche sexuelle Ausrichtung ausdrücken. Das Wort „gay“ wird mit 0.73 Prozent (Stand 22.11.2017) als „toxic“ gewertet. In den Fällen von Wikipedia und der New York Times ist dieser Kontext zwar auch vorhanden, besonders aber in Hinsicht auf Social Media kommt diesem noch einmal eine wesentlich größere Bedeutung zu. Es besteht die Gefahr, die persönliche Freiheit einzuschränken.

Ein anderer Aspekt leitet sich ebenfalls aus dem Kontext ab, nur entgegen gesetzt, andere Wörter können nämlich im jeweiligen Kontext beleidigend oder als „Troll“ wirken, obwohl sie im normalen Sprachgebrauch nicht negativ konnotiert sind. Ein weiteres Problem könnte sein, dass viele immer einen Weg finden werden, ihren Beitrag zu senden. Sei es, indem sie Synonyme verwenden oder sich immer neue Möglichkeiten erdenken. Es ist auch bezeichnend, dass Perspective nicht in Googles eigenen Projekten zum Einsatz kommt. YouTube hat teilweise sehr große Probleme in den Kommentarbereichen unterhalb von Videos, trotzdem kommt Perspective an dieser Stelle noch nicht zum Einsatz.

Ich glaube, dass die Möglichkeiten mit intelligenten Algorithmen zu arbeiten, in den nächsten Jahren wichtig wird und eine gute Waffe im Kampf gegen Hasskommentare und Trolls sein kann. Jedoch glaube ich, dass Perspective in den nächsten Jahren noch nicht in der Lage sein wird, den Kontext eines Kommentars für eine effektive Auswertung zu verwenden. Jedoch kann ich mir vorstellen, dass ein Einsatz in Kombination mit Moderatoren gelingen kann.

D. Politik

Dem Thema Diskussionskultur im Internet kann auch die deutsche Politik sich nicht entziehen. Die wichtigsten Aspekte sind hier besonders für den Wahlkampf „Fake News“ und „Social Bots“, aber auch Hasskommentare generell sind ein großes Thema.

Der Vergleich zum US Wahlkampf liegt nahe, da in diesem wie schon weiter oben beschrieben, jene Techniken besonderes stark zum Einsatz kamen. Mit diesem Thema hat sich das Oxford Internet Institute der Universität von Oxford im Rahmen einer kleinen Studie[25] genauer beschäftigt. Dafür wurden insgesamt über eine Million Tweets untersucht. Zum Einen fanden sie heraus, dass im Vergleich zum US Wahlkampf, deutlich weniger „Fake News“ im Umlauf sind. Das liegt unter anderem daran, dass sich die Nutzer wesentlich mehr auf die etablierten Medien verlassen.

Auch der Einsatz von „Social Bots“ sei deutlich geringer. Sie machten 7,6 Prozent der untersuchten Twitter-Nachrichten aus. Das liegt vermutlich auch daran, dass sich die Parteien im Herbst 2016 gemeinschaftlich gegen „Social Bots“ ausgesprochen haben[26]. Der größte Einsatz von „Social Bots“ wurde im Zusammenhang mit der AfD festgestellt, welche sich erst zwei Tage nach den anderen Parteien ebenfalls für den Verzicht von „Social Bots“ ausgesprochen hat.

Aber auch über den Wahlkampf hinaus, versucht die Regierung der Bundesrepublik Einfluss auf die Diskussion im Internet zu nehmen. Ein Gesetzesentwurf, welcher vom Kabinett beschlossen wurde sieht vor, dass Betreiber sozialer Netzwerke ab zwei Millionen Nutzer verpflichtet sind, „offensichtlich strafbare Inhalte wie Verleumdung oder Volksverhetzung innerhalb von 24 Stunden nach Eingang der Beschwerde zu löschen oder zu sperren“ [27]. Bei Missachtung drohen Strafen in Höhe von bis zu 50 Millionen Euro.

Der Entwurf steht jedoch von verschiedenen Seiten in der Kritik (siehe [27]). Als Hauptproblem wird dabei angesehen, dass die Meinungsfreiheit eingeschränkt werden könnte. Denn mit dem Gesetzesentwurf liegt die Verantwortlichkeit bei den Sozialen Netzwerken. Diese waren schon vorher verpflichtet, rechtswidrige Inhalte zu entfernen. Befürchtet wird, dass diese aus Angst vor Strafe alle Inhalte unter Verdacht vorsorglich löschen.

Persönlich halte ich den Einfluss der Politik für begrenzt. Das liegt unter anderem daran, dass sie selber Teil des Problems ist. Besonders bei politisch rechten Parteien ist die Ablehnung der beschriebenen Phänomene nicht besonders groß. Es kann teilweise beobachtet werden, dass diese zumindest im Zusammenhang mit dem eigenen Wahlkampf öfter eingesetzt werden, als von anderen Parteien. Für diese Parteien geht jeder Einschnitt auf Kosten der Meinungsfreiheit. Hinzu kommt, dass der Einfluss der deutschen Politik auf die, vor allem in den USA ansässigen, Sozialen Netzwerke eher beschränkt ist. Jedoch ist auch zu erkennen, dass der deutschen Politik das Problem bewusst ist. So ist die Absprache über den Verzicht

auf „Social Bots“ auf jeden Fall positiv zu bewerten. Auch der Gesetzesentwurf für Soziale Netzwerke kann als positiver Impuls angesehen werden. Besonders, wenn sich dieser in die EU fortsetzt und es so eine europaweite, einheitliche Gesetzgebung gäbe, könnte das starke Auswirkung haben. Letztlich soll die Politik die Rahmenbringungen vorgeben und es wird sich zeigen, ob es ihr gelingt die Betreiber zu erhöhter Aufmerksamkeit zu zwingen.

E. Anti-Hass-Design

Eine weitere Möglichkeit die Diskussionskultur zu verbessern und „Toxic Communitys“ vor der Entstehung zu verhindern, ist zu versuchen mit einem Design Ansatz alle möglichen Reibungspunkte im Vorfeld zu erkennen und sie zu minimieren oder sogar ganz zu entfernen.

Besonders im Bereich von Online-Spielen sind „Toxic Communitys“ verbreitet, in diesem Bereich aber auch sehr differenziert. Es gibt Communitys, welche nur geringe Probleme mit dieser Thematik haben, andere besitzen einen überwältigend schlechten Ruf. Das Medium ist besonders gefährdet, bietet jedoch auch viele Möglichkeiten.

Ein Beispiel für ein Projekt, welches versucht hat schon im Konzept (siehe [28]) zu verhindern, dass sich Nutzer gegenseitig angehen, ist der Heldenshooter Overwatch. Bereits mit der Comic-Grafik wurde gezielt auf eine helle fröhliche Erfahrung gesetzt. Außerdem gibt es nicht die Möglichkeit in Deathmatches gegen einander anzutreten, sondern ausschließlich Spielmodi, in welchen es darum geht, als Team zu agieren und Aufgaben zu erfüllen. Dies ist auch der Grund, weshalb auf eine „Kill/Death“-Statistik verzichtet wurde. Generell gibt es kein Spieler-Ranking und es wurde darauf verzichtet die Spieler miteinander zu vergleichen. Die einzige Form der Anerkennung von Leistungen sind Auszeichnungen, welche Spieler bekommen, wenn diese ihre Rolle besonders gut erfüllt haben.

Trotz dieser Bemühungen gab der Chefentwickler Jeff Kaplan im Herbst 2017 in einem Entwicklerupdate auf YouTube bekannt[29], dass Entwickler-Ressourcen von der Entwicklung neuer Features abgezogen werden mussten, damit diese eingesetzt werden konnten, um etwas gegen die toxische Community zu unternehmen. Dies zeigt, dass die Bemühungen umsonst waren und trotz des großen Aufwands, ein Design zu entwickeln, welches das Miteinander fördert, gilt die Community mittlerweile als „toxic“.

Overwatch ist ein gutes Beispiel dafür, dass auch ein gutes Design nicht verhindern kann, dass eine Community „toxic“ werden kann. Dabei kann ein solches Design zwar generell dazu beitragen, das Klima einer Gemeinschaft zu verbessern, jedoch ist es nicht möglich, Verhaltensweisen, welche im Internet, besonders im Bereich von Multiplayer-Spielen, üblich sind, nur mit einem solchen Design und Moderatoren auszusperren.

VI. ZUSAMMENFASSUNG UND FAZIT

Die fünf vorgestellten Maßnahmen besitzen alle unterschiedliche Vor- und Nachteile. In der Tabelle I werden diese nach ihren Eigenschaften aufgelistet und noch nach drei Kriterien bewertet. Im Folgenden wird die Bewertung genauer begründet und die Methoden auf die Einsatzmöglichkeit hinsichtlich der fünf Ausprägungen Shitstorm, Fake News, Filterblase, Social Bots und Toxic Communitys dargestellt.

Das erste Kriterium ist **Präventiv** und beschreibt, wie geeignet die Methode ist, um eine Eskalation noch vor der Entstehung zu verhindern. Der nächste Punkt beschreibt den **Wirkungsgrad**. Dieser beschreibt, wie gut die Methode die zugewiesene Aufgabe erfüllt. Der letzte Punkt **Akzeptanz** bewertet, ob die Nutzer sich der entsprechenden Vorgehensweise fügen oder diese ablehnen.

Counterspeech hat sich als sehr ineffektiv herausgestellt.

Tabelle I
AUSWERTUNG

Maßnahme	Präventiv	Wirkungsgrad	Akzeptanz
Counterspeech	nein	eher gering	eher gering
Moderatoren	teilweise	mittel	teilweise
Perspektive	ja	stark	teilweise
Politik	-	-	ja
Design	ja	eher gering	ja

Es verhindert nicht, dass Hass entstehen kann, viel eher kann es dazu führen, dass eine Diskussion von neuem aufflammt. Der eigentliche Nutzen steht im Gegenteil gegen Hass. Ein Vorteil besteht darin, dass Counterspeech bei allen Ausprägungen eingesetzt werden kann, um auf deren Inkorrektheit hinzuweisen. Auch wenn die Wirkungsmöglichkeit eher gering sind.

Moderatoren sind ein essentieller Bestandteil der Online Diskussion. Ebenso wie Counterspeech, können sie gegen die fünf ausgewählten Ausprägungen wirken. Dabei bedienen sie sich unter anderem auch dieser. Hinzu kommen jedoch die weiteren Möglichkeiten, mit welchen sie unter anderem unbelehrbare Nutzer von einer Diskussion ausschließen können. Als großer Nachteil ist jedoch zu sehen, dass es schwer bis unmöglich ist, die immensen Mengen zu kontrollieren. Dies bedeute einen immensen Aufwand und führt dazu, dass meist nicht schnell genug reagiert werden kann, um etwa einen Shitstorm noch im Keim zu ersticken. Außerdem kann eine Aktion eines Moderators auch dazu führen, dass eine Diskussion sich noch weiter erhitzt, da die Nutzer dies als Angriff auf ihre Freiheit sehen.

Die Künstliche Intelligenz Perspektive ist ein sehr innovatives Projekt, welches versucht einen Moderator mit einem Programm zu ersetzen. Der aktuelle Entwicklungseinblick zeigt das immense Potenzial. Da schon bei der Erstellung von Kommentaren überprüft werden kann, ob sich diese an die Richtlinien halten, ist der Wirkungsgrad enorm. Jedoch zeigt es auch noch Schwächen. Der Interpretationsgrad und andere Lücken sind nicht einfach zu lösende Probleme. Ebenso kann

Perspektive z.B. noch keine Social Bots oder Fake News erkennen. Jedoch ist es durchaus möglich, dass die K.I. auch solche Aufgaben in Zukunft lösen kann.

Bei der Politik kann keine eindeutige Bewertung vorgenommen werden. Es wäre zwar möglich, dass jene vorsorgend Gesetze und Grundlagen schaffen, meist jedoch kommt es durch die rasante Entwicklung und langwierige Verfahren in der Legislative, erst spät zu solchen Regelungen. Der Wirkungsgrad ist ebenfalls schwierig zu bewerten, da eher Rahmenbedingungen geschaffen werden. Dabei besitzen diese jedoch ein großes Potenzial. So sind sie die einzigen, welche effektiv etwas gegen Filterblasen und Falschmeldungen unternehmen könnten. Es wäre z.B. möglich Ernsthaftige Falschmeldungen, hinsichtlich Verleumdung, härter zu bestrafen. Eine anderer möglicher Eingriff wäre, große Plattformbetreiber und Suchmaschinen zu einer gewissen Form von Diversität zu verpflichten. Diese müssten dann z.B. mehr alternative Beiträge anzeigen, um so die Echokammer aufzuweichen.

Der Ansatz vorbeugend ein Design zu entwickeln, welches eine angenehme Diskussionskultur schafft, ist besonderes in der Prävention sehr stark. Jedoch zeigt das Beispiel Overwatch, dass es noch keine effektive Umsetzung gibt und es sich so noch im Konzeptstadium befindet. Dabei kann auch davon ausgegangen werden, dass es einen Konflikt zwischen dem Wirkungsgrad und der Akzeptanz gibt, je stärker die Methode in das Erleben einschneidet. Auch hier können Fake News und Social Bots nicht ausgeschlossen werden, jedoch könnte ein entsprechendes Design durchaus dazu beitragen, gegen die Filterblase zu wirken.

Abschließend kann festgestellt werden, dass es viele unterschiedliche Ausprägungen gibt, welche die Online-Diskussionskultur schädigen. Um diesen entgegen zu wirken, gibt es die unterschiedlichsten Herangehensweisen. Jedoch löst keine ein Problem vollständig, geschweige denn, wenn mehrere zusammenwirken. Das wohl größte Potenzial bietet eine K.I. gestützte Methode, diese steht allerdings noch am Anfang. In Zukunft wird es jedoch wahrscheinlich, dass eine solche Lösungen den Diskurs im Internet steuern wird.

LITERATURVERZEICHNIS

[1] Liriam Sponholz. *Hate Speech*. Wiesbaden: Springer Fachmedien Wiesbaden, 2018. ISBN: 978-3-658-15077-8. DOI: 10.1007/978-3-658-15077-8_2. URL: <https://link.springer.com/book/10.1007/978-3-658-15077-8#about> (siehe Seite 1, 4).

[2] B. Batinic. *Internet für Psychologen*. Hogrefe Verlag, 2000. ISBN: 9783840912269. URL: <https://books.google.de/books?id=Ob4jLyrQZ4Kc> (siehe Seite 1).

[3] Tamara Shepherd, Alison Harvey, Tim Jordan u. a. "Histories of Hating". In: *Social Media + Society* 1.2 (2015), Seite 2056305115603997. DOI: 10.1177/2056305115603997. eprint: <https://doi.org/10.1177/2056305115603997>. URL: <https://doi.org/10.1177/2056305115603997> (siehe Seite 1).

[4] L. Steinke. *Bedienungsanleitung für den Shitstorm: Wie gute Kommunikation die Wut der Masse bricht*. essentials. Springer Fachmedien Wiesbaden, 2014. ISBN: 9783658055882. URL: <https://books.google.de/books?id=pUQIBAAQBAJ> (siehe Seite 2).

[5] Dan Wootton. "Sheeran twitter quitter". In: *The Sun* (Juli 2017). URL: <https://www.thesun.co.uk/tvandshowbiz/3939544/ed-sheeran-forced-to-quit-twitter-after-revealing-vile-comments-bring-him-down/> (besucht am 25. 10. 2017) (siehe Seite 2).

[6] Duden. *Shitstorm, der*. URL: <https://www.duden.de/rechtschreibung/Shitstorm> (besucht am 30. 11. 2017) (siehe Seite 2).

[7] Matthias Schulz. "Penny: Facebook-Streit um Schoko-"Zipfelmann"". In: *Spiegel Online* (2015). URL: <http://www.spiegel.de/wirtschaft/service/schoko-zipfelmann-von-penny-kritik-auf-facebook-a-1066710.html> (besucht am 27. 10. 2017) (siehe Seite 2).

[8] Zeit. *Was sind Fake News?* URL: <http://www.zeit.de/thema/fake-news> (besucht am 27. 10. 2017) (siehe Seite 2).

[9] Victoria L. Rubin, Yimin Chen und Niall J. Conroy. "Deception detection for news: Three types of fakes". In: *Proceedings of the Association for Information Science and Technology* 52.1 (2015), Seiten 1–4. ISSN: 2373-9231. DOI: 10.1002/pr2.2015.145052010083. URL: <http://dx.doi.org/10.1002/pr2.2015.145052010083> (siehe Seite 2).

[10] SZ. "Asylbewerber-Angriff erfunden". In: *Sächsische Zeitung* (12.05.2015). URL: <http://www.sz-online.de/nachrichten/asylbewerber-angriff-erfunden-3101651.html> (besucht am 05. 11. 2017) (siehe Seite 2).

[11] Wolfgang Wichmann. "Kampf gegen "Online-Hetzjagd"". In: *tagesschau* (10.07.2017). URL: <http://faktenfinder.tagesschau.de/inland/gzwanzig-143.html> (besucht am 28. 10. 2017) (siehe Seite 2).

[12] Eva Ihnenfeldt. "Was ist eigentlich... eine „Filterblase“? Sachen, die man wissen sollte 80". In: *steadynews.de* (2017). URL: <https://steadynews.de/marketing/was-ist-eigentlich-eine-filterblase-sachen-die-man-wissen-sollte> (besucht am 18. 11. 2017) (siehe Seite 3).

[13] Heinrich-Böll-Stiftung. "Filter Bubble – Echokammer – Fake News". In: *Heinrich Böll Stiftung* (2017). URL: https://www.boell.de/de/2017/02/09/social-bots?dimension1=ds_digitaler_wahlkampf (besucht am 05. 11. 2017) (siehe Seite 3).

[14] Heinrich-Böll-Stiftung. "Social Bots". In: *Heinrich Böll Stiftung* (2017). URL: https://www.boell.de/de/2017/02/09/social-bots?dimension1=ds_digitaler_wahlkampf (besucht am 05. 11. 2017) (siehe Seite 3).

[15] By Adam Rosenberg. "What changed for 'Call of Duty' after all the YouTube hate? Nothing". In: *mashable* (2016). URL: <http://mashable.com/2016/06/22/call-of-duty-infinite-warfare-eric-hirshberg-interview/?#5YqZkrqOpagp> (besucht am 20. 11. 2017) (siehe Seite 4).

[16] Giuseppe Nelva. "NPD: Call of Duty Tops December US Sales; Final Fantasy XV Ranks Second but Wins on PS4". In: *DualSHOCKERS* (2017). URL: <http://www.dualshockers.com/npd-call-duty-tops-december-us-sales-final-fantasy-xv-ranks-second-wins-ps4/> (besucht am 20. 11. 2017) (siehe Seite 4).

[17] Katrin Gottschalk. "Wer Sexismus anprangert, wird mit Vergewaltigung bedroht". In: *Spiegel Online* (2014). URL: <http://www.spiegel.de/netzwelt/netzpolitik/anita-sarkeesian-feministische-videospielkritik-und-morddrohungen-a-988906.html> (besucht am 20. 11. 2017) (siehe Seite 4).

[18] Charles Randall. "Open Letter". In: *Twitter* (2017). URL: <https://twitter.com/charlesrandall/status/911987526541430784> (besucht am 20. 11. 2017) (siehe Seite 4).

[19] Astrid Herbold. "Argumente sind kein Allheilmittel". In: *Zeit Online* (2016). URL: <http://www.zeit.de/digital/internet/2016-02/hatespeech-counterspeech-facebook-forschung> (besucht am 28. 11. 2017) (siehe Seite 4, 5).

[20] eco. *Jeder Dritte ist schon rassistischen Hassbotschaften im Internet begegnet*. Feb. 2016. URL: <https://politik-recht.eco.de/2016/pressemitteilung/ecomfrage-jeder-dritte-ist-schon-rassistischen-hassbotschaften-im-internet-begegnet-2.html> (besucht am 26. 11. 2017) (siehe Seite 4).

[21] Jochen Mai. "Community-Manager: Berufsbild und Tipps". In: *karriere bibel* (2012). URL: <https://karrierebibel.de/community-manager/> (besucht am 28. 11. 2017) (siehe Seite 5).

[22] New York Times. *The Times is Partnering with Jigsaw to Expand Comment Capabilities*. 2016. URL: <https://www.nytc.com/the-times-is-partnering-with-jigsaw-to-expand-comment-capabilities/> (besucht am 28. 11. 2017) (siehe Seite 5).

[23] Jigsaw. *Perspective*. 2017. URL: <https://www.perspectiveapi.com/#/> (besucht am 22. 11. 2017) (siehe Seite 5).

[24] Violet Blue. "Google's comment-ranking system will be a hit with the alt-right". In: *engadget* (2017). URL: <https://www.engadget.com/2017/09/01/google-perspective-comment-ranking-system/> (besucht am 22. 11. 2017) (siehe Seite 6).

[25] Philip N. Howard Lisa-Maria Neudert Bence Kollanyi. "Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter?". In: *Oxford Internet Institute, University of Oxford* (2017). URL: <http://comprop.oii.ox.ac.uk/publishing/junk-news-and-bots-during-the-german-parliamentary-election-what-are-german-voters-sharing-over-twitter/> (besucht am 23. 11. 2017) (siehe Seite 6).

[26] Martin Mair. "Kampf den Meinungsmaschinen". In: *Tagesschau* (2017). URL: <http://www.tagesschau.de/inland/social-bots-wahlkampf-101.html> (besucht am 23. 11. 2017) (siehe Seite 6).

[27] Jörg Seisselberg. "Soziale Netzwerke in der Pflicht". In: *Tagesschau* (2017). URL: <https://www.tagesschau.de/inland/hasskommentare-103.html> (besucht am 23. 11. 2017) (siehe Seite 6).

[28] Daniel Starkey. "Overwatch: A Laid-Back Shooter Without the Crappy Attitudes". In: *wired* (2016). URL: <https://www.wired.com/2016/05/overwatch-first-impressions/> (besucht am 20. 11. 2017) (siehe Seite 7).

[29] Jeff Kaplan. *Seid freundlich, spielt fair! | Entwicklerupdate*. 2017. URL: <https://www.youtube.com/watch?v=IhSF6hXN7PU&feature=youtu.be> (besucht am 20. 11. 2017) (siehe Seite 7).