



# MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„Wie gehen Unternehmen mit Harmful Online  
Communication um?

Eine detaillierte länderübergreifende Inhaltsanalyse von  
Unternehmensrichtlinien.“

verfasst von / submitted by

Talin Schrattenholzer, Bakk. phil.

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of  
Magistra der Philosophie (Mag. phil.)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt /  
Degree programme code as it appears on  
the student record sheet:

UA 066 841

Studienrichtung lt. Studienblatt /  
Degree programme as it appears on  
the student record sheet:

Publizistik- und Kommunikationswissenschaft

Betreut von / Supervisor:

Univ.-Prof. Dr. Sabine Einwiller



# Inhaltsverzeichnis

I. Einleitung .....	1
II. Harmful Online Communication.....	3
2.1. Begriffsbestimmungen .....	3
2.2. Akteure und ihre Motive .....	6
2.3. Faktoren, die Hass im Netz begünstigen.....	9
2.3.1. Anonymität.....	9
2.3.2. Rechtsfreier Raum .....	10
2.3.3. Nasty-Effect .....	11
2.3.4. Echokammern und Filterblasen .....	12
2.4. Auswirkungen von Harmful Online Communication .....	14
2.5. Gesetzliche Grundlagen .....	17
2.5.1. Das Recht auf freie Meinungsäußerung .....	17
2.5.2. Gesetzliche Lage in Österreich .....	21
2.5.3. Gesetzliche Lage in Deutschland .....	24
2.5.4. Gesetzliche Lage in den Vereinigten Staaten .....	27
2.5.5. Gesetzliche Lage in Großbritannien .....	28
III. Die Verantwortung von Unternehmen und ihre Maßnahmen gegen HOC.....	31
3.1. Lösungsvorschläge für den Umgang mit HOC.....	33
3.1.1. Verhaltensrichtlinien und Forenregeln .....	33
3.1.2. Registrierung auf Online-Plattformen.....	34
3.1.3. Moderation der Online-Plattformen.....	35
3.1.4. Umgang mit HOC-Inhalten .....	37
IV. Wie Betroffene gegen HOC vorgehen können .....	39
4.1. Blockieren und Melden .....	39
4.2. Counter Speech .....	40
4.3. Rechtliche Schritte .....	42
V. Empirie.....	43
5.1. Methode .....	43
5.1.1. Methodische Vorgehensweise .....	44
5.1.2. Stichprobe .....	44
3.1.3. Kategorienbildung .....	49

5.1.4. Codebuch .....	50
5.1.5. Pretest .....	53
5.1.6. Reliabilitätstest .....	54
5.2. Ergebnisse der Inhaltsanalyse.....	57
5.2.1. Inhalte der Richtlinien.....	59
5.2.2. Maßnahmen gegen HOC .....	60
5.2.3. Länderspezifische Unterschiede .....	67
5.3. Diskussion der Ergebnisse .....	78
VI. Fazit.....	83
VII. Quellenverzeichnis .....	85
VIII. Anhang .....	95
8.1. Codebuch .....	95
8.2. Reliabilitätstest .....	107
Abstract Deutsch.....	108
Abstract English .....	109

# Abkürzungsverzeichnis

CDA Communication Decency Act

ECG E-Commerce-Gesetz

ECRI Europäische Kommission gegen Rassismus und Intoleranz

EGMR Europäischer Gerichtshof für Menschenrechte

EMRK Europäische Menschenrechtskonvention

GG Grundgesetz

HOC Harmful Online Communication

MCA Malicious Communications Act

MedienG Mediengesetz

NetzDG Netzwerkdurchsetzungsgesetz

OGH Oberster Gerichtshof

POA Public Order Act

SNS Social Network Sites

StGB Strafgesetzbuch

StGG Staatsgrundgesetz

VG Verbotsgesetz



# Abbildungsverzeichnis

Abbildung 1: Einstellungen zur freien Meinungsäußerung (Auswahl) .....	20
Abbildung 2: Einstellungen zur Zensur von Aussagen, die gegenüber Minderheiten anstößig sind.....	21
Abbildung 3: Herkunft der untersuchten Richtlinien (in %) .....	57
Abbildung 4: Anzahl der Dokumente aus den jeweiligen Kategorien .....	58
Abbildung 5: Anteil der analysierten Richtlinien nach Art (in %).....	59
Abbildung 6: Einstufung der unterschiedlichen Inhalte als HOC laut den Richtlinien .....	60
Abbildung 7: Anteil der Richtlinien, in denen erwähnt wird, wie Userinnen und User interagieren sollen (in %) .....	61
Abbildung 8: Erwünschtes Verhalten, das in den Richtlinien von den Userinnen und Usern erwartet wird.....	62
Abbildung 9: Handlungsmöglichkeiten der Userinnen und User in Bezug auf HOC .....	63
Abbildung 10: Anteil der Richtlinien, die eine detaillierte Anleitung zum Melden von HOC beinhalten (in %) .....	64
Abbildung 11: Anteil der Richtlinien in denen erwähnt wird, wie die Unternehmen mit den gemeldeten Inhalten umgehen (in %) .....	65
Abbildung 12: Gründe für ein Verbot von HOC .....	66
Abbildung 13: Konsequenzen beim Verstoß gegen die Richtlinien .....	67
Abbildung 14: Anteil der unterschiedlichen Richtlinienarten nach Land (in %) .....	68
Abbildung 15: Einstufung der unterschiedlichen Inhalte als HOC laut den Richtlinien nach Land (in %) .....	69
Abbildung 16: Anteil der Erwähnungen von Kinder- und Jugendschutzbestimmungen nach Land (in %) .....	70
Abbildung 17: Anteil der Richtlinien, in denen erwähnt wird, wie Userinnen und User interagieren sollen, nach Land (in %) .....	71
Abbildung 18: Länderspezifische Unterschiede beim von den Userinnen und Usern in den Richtlinien erwünschten Verhalten (in %) .....	72
Abbildung 19: Länderspezifische Unterschiede in den Handlungsmöglichkeiten der Userinnen und User in Bezug auf HOC (in %) .....	73
Abbildung 20: Anteil der Richtlinien, die eine detaillierte Anleitung zum Melden von HOC beinhalten, nach Land (in %) .....	74
Abbildung 21: Anteil der Richtlinien, in denen erwähnt wird, wie die Unternehmen mit den gemeldeten Inhalten umgehen, nach Land (in %) .....	75

Abbildung 22: Länderspezifische Unterschiede hinsichtlich der Gründe für ein Verbot von HOC .....	76
Abbildung 23: Länderspezifische Unterschiede hinsichtlich der Konsequenzen beim Verstoß gegen die Richtlinien.....	77



# Tabellenverzeichnis

Tabelle 1: Ausgewählte Unternehmen aus Österreich .....	45
Tabelle 2: Ausgewählte Unternehmen aus Deutschland .....	46
Tabelle 3: Ausgewählte Unternehmen aus den Vereinigten Staaten .....	47
Tabelle 4: Ausgewählte Unternehmen aus Großbritannien .....	48
Tabelle 5: Anzahl und Art der analysierten Richtlinien pro Land .....	49
Tabelle 6: Kategorien 1-5 .....	50
Tabelle 7: Anzahl der codierten Unternehmensrichtlinien für den Pretest .....	54
Tabelle 8: Anzahl der codierten Unternehmensrichtlinien für den Reliabilitätstest .....	55



# I. Einleitung

Hass in sozialen Medien, in Online-Diskussionsforen oder Onlinemedien ist in den vergangenen Jahren zu einem schwerwiegenden gesellschaftlichen Problem geworden. Hass im Netz äußert sich meist in Form von diskriminierenden, menschenverachtenden oder aggressiven Kommentaren. Derartige Äußerungen, auch „Hate Speech“ und von Einwiller und Kim „Harmful Online Communication“ genannt, richten sich in den meisten Fällen gegen einzelne Personen, soziale Gruppen oder Unternehmen (Einwiller & Kim, 2018).

Um Hass im Netz zu bekämpfen, ist es wichtig, die Öffentlichkeit auf das Phänomen aufmerksam zu machen und darzulegen, wie destruktiv – also „harmful“ – diese Vorgänge für ein gelingendes Zusammenleben und eine gute Gesprächskultur sind. Das 2016 in Österreich gegründete „No Hate Speech“-Komitee operiert beispielsweise in diese Richtung. Es basiert auf der 2013 ins Leben gerufenen „No Hate Speech“-Kampagne des Europarates. Ziel dieser Initiative ist es, Menschenrechte zu fördern, von Hate Speech Betroffene zu unterstützen und die Öffentlichkeit für das Thema Hass im Netz zu sensibilisieren, um somit der Akzeptanz von Hassreden entgegenzuwirken (Nationales Komitee No Hate Speech Österreich, 2018).

Neben den nationalen und internationalen Institutionen verpflichten sich mittlerweile auch Unternehmen wie Facebook, Twitter, YouTube oder Microsoft zur Bekämpfung illegaler Hassrede im Internet. In Zusammenarbeit mit der Europäischen Kommission wurde ein Verhaltenskodex entwickelt, der darauf abzielt, die Verbreitung von unerwünschten Hassbotschaften auf Online-Plattformen zu unterbinden. Darin verpflichten sich die Unternehmen unter anderem, rechtswidrige und hasserfüllte Äußerungen innerhalb von 24 Stunden nach deren Bekanntwerden zu entfernen. Zusätzlich wurde vereinbart, dass die Unternehmen durch ihre Richtlinien kommunizieren müssen, dass gewisse Inhalte wie etwa Hassbotschaften auf ihren Online-Plattformen unzulässig sind und daher nicht toleriert werden (Europäische Kommission, 2016).

Maßnahmen wie die der Europäischen Kommission legen den Grundstein für eine aufgeklärte, sensibilisierte Gesellschaft. Denn es liegt in der Verantwortung von Unternehmen, die ihren Nutzerinnen und Nutzern Plattformen zur öffentlichen Diskussion zur Verfügung stellen, gegen rechtswidrige Inhalte vorzugehen. Ob auch weitere

Unternehmen dem Beispiel von Facebook, Twitter & Co folgen, soll in der vorliegenden Arbeit erforscht und dargelegt werden. Konkret soll herausgefunden werden, wie Unternehmen, die Kommentare und Diskussionen auf ihren Online-Plattformen zulassen, mit Harmful Online Communication (HOC) umgehen. Hierfür werden unterschiedliche Unternehmensrichtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien inhaltlich analysiert.

Auf Grund dieses Erkenntnisinteresses sind folgende Forschungsfragen erstellt worden:

**FF1:** Welche Inhalte zählen laut der Unternehmensrichtlinien zu HOC?

**FF2:** Welche Maßnahmen werden von den Unternehmen getroffen, um gegen HOC vorzugehen?

**FF3:** Inwiefern können länderspezifische Unterschiede in Bezug auf HOC festgestellt werden?

## II. Harmful Online Communication

### 2.1. Begriffsbestimmungen

Wie bereits erwähnt, richtet sich der digitale Hass in den meisten Fällen gegen bestimmte Personen oder Personengruppen. Mit diversen Kraftausdrücken und einer überwiegend aggressiven Sprache werden Personen unter anderem aufgrund ihrer Hautfarbe, Religion, Nationalität, einer Behinderung, des Geschlechts oder der sexuellen Orientierung herabgesetzt und verunglimpft (Meibauer, 2013). Es handelt sich hierbei um eine sehr respektlose und gewalttätige Art der Kommunikation, die darauf abzielt, Personen zu demütigen und zu verletzen (Grimm, 2016).

Zwar gibt es bisher keine allgemeingültige einheitliche Definition des Begriffs „Hassrede“, allerdings weisen die unterschiedlichen Definitionen einige Gemeinsamkeiten auf. In der Definition des Ministerkomitees des Europarates aus dem Jahr 1997 wird beispielsweise ausdrücklich die rassistische Diskriminierung von bestimmten Personengruppen angesprochen. Definiert wird der Begriff „Hassrede“ in der Empfehlung wie folgt:

„Jegliche Ausdrucksformen, welche Rassenhass, Fremdenfeindlichkeit, Antisemitismus oder andere Formen von Hass, die auf Intoleranz gründen, propagieren, dazu anstiften, sie fördern oder rechtfertigen, einschliesslich der Intoleranz, die sich in Form eines aggressiven Nationalismus und Ethnozentrismus, einer Diskriminierung und Feindseligkeit gegenüber Minderheiten, Einwanderern und der Einwanderung entstammenden Personen ausdrücken.“  
(Europarat/Ministerkomitee, 1997, S. 2)

Auch die „Allgemeine Politik-Empfehlung Nr. 15 über die Bekämpfung von Hassrede“ der Europäischen Kommission gegen Rassismus und Intoleranz (ECRI) geht explizit auf die rassistisch motivierte Hassrede ein. Allerdings werden in dieser Definition auch weitere Formen der Diskriminierung miteinbezogen, da hier auch von Hass und Herabwürdigungen gegen Personen oder Personengruppen unter anderem aufgrund ihres Alters, einer Behinderung, des Geschlechts oder der sexuellen Orientierung gesprochen wird. In der Empfehlung hält die Kommission fest, dass unter Hassrede:

„das Befürworten und Fördern von oder Aufstacheln zu jeglicher Form von Verunglimpfung, Hass oder Herabwürdigung einer Person oder Personengruppe zu verstehen ist, ebenso wie jegliche Belästigung, Beleidigung, negative Stereotypisierung, Stigmatisierung oder Bedrohung einer Person oder Personengruppe und die Rechtfertigung der genannten Äußerungen, die aufgrund der „Rasse“, Hautfarbe, Abstammung, nationalen oder ethnischen Herkunft, des Alters, einer Behinderung, der Sprache, der Religion oder der Überzeugung, des biologischen oder sozialen Geschlechts, der Geschlechtsidentität, sexuellen Orientierung oder anderer persönlicher Eigenschaften und Statusmerkmale getätigt werden.“ (Europäische Kommission gegen Rassismus und Intoleranz, 2016, S. 3)

Die eben zitierten Definitionen haben eines gemeinsam: sie zielen allesamt auf eine Diskriminierung von Personen aufgrund eines gemeinsamen Merkmales (Herkunft, Religion, sexuelle Orientierung etc.) ab. Hate Speech ist also zum einen dadurch gekennzeichnet, dass Menschen aufgrund eines gemeinsamen Merkmales kategorisiert werden (Sponholz, 2018) und aufgrund dieses Merkmales mit stereotypisierten, hasserfüllten Äußerungen (Warner & Hirschberg, 2012) konfrontiert werden. Warner und Hirschberg (2012) merken dabei zurecht an, dass stereotypisierte Äußerungen nicht zwangsläufig explizit hasserfüllt und mit Kraftausdrücken versehen sein müssen, um als Hassrede zu gelten. Stereotypisierte Hassbotschaften können unterschiedliche Formen annehmen und durchaus auch subtil formuliert und aus (auf den ersten Blick) harmlosen Wörtern bestehen. Bei diesen Äußerungen komme es vielmehr auf die Absicht der Autorin oder des Autors an, die letztendlich ausschlaggebend für die Beurteilung ist, ob eine Äußerung unter Hassrede fällt oder nicht.

In vielen Definitionen wird auch auf die für Hassreden typische Sprache eingegangen, die überwiegend als beleidigend und missbräuchlich (Warner & Hirschberg, 2012), abfällig und demütigend (Davidson, Warmsley, Macy, & Weber, 2017), schädlich (Gelber, 2019) oder gar gefährlich (Cohen-Almagor, 2017) beschrieben wird.

In der vorliegenden Arbeit wird der Begriff „Harmful Online Communication“ (HOC) gegenüber „Hate Speech“ bevorzugt, da er sowohl den Aspekt der Diskriminierung als auch die Auswirkungen der hasserfüllten Sprache aufweist. Einwiller und Kim definieren HOC wie folgt:

„Harmful online communication means ways of expression in online environments containing aggressive and destructive diction that violate social norms and aim at

harming the dignity or safety of the attacked target, which can be a person, a social group or an organization.” (Einwiller & Kim, 2018, S. 2)

Einwiller und Kim merken in ihrer Arbeit an, dass der Begriff HOC das Phänomen „Hass im Netz“ umfassender beschreibt als Hassrede, da er einerseits die Form des Ausdrucks (aggressiv, hasserfüllt oder destruktiv) sowie deren potenzielle Wirkung (schädlich oder verletzend) auf die Betroffenen beinhaltet. Dass das Schädigungspotential von Hass im Internet so groß ist, liege an der Langlebigkeit und Viralität von Hassbotschaften (Einwiller & Kim, 2018). Viele (oftmals auch anonym oder unter einem Pseudonym verfasste) Hassbotschaften erreichen durch das Posten, Liken und Teilen auf sozialen Netzwerken ein breites Publikum. Diese Inhalte aus dem Internet zu entfernen und die Schuldigen zur Rechenschaft zu ziehen, gestaltet sich als äußerst schwierig.

## 2.2. Akteure und ihre Motive

Nun stellt sich die Frage, ob Personen, die sich einer derartig aggressiven und destruktiven Sprache bedienen, typologisiert werden können. Ingrid Brodnig unterscheidet in ihrem Buch „Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können“ (2016) zwei Typen von Userinnen und Usern, die aus unterschiedlichen Gründen Hass im Netz verbreiten. Da gibt es einerseits jene Personen, auch „Trolle“ genannt, die gezielt „Trolling“ im Internet betreiben. Laut Ingrid Brodnig kommt der Begriff „Trolling“ aus der Anglersprache und bezeichnet eine bestimmte Vorgehensweise von Fischern, bei der Fische durch das Auswerfen von Ködern gelockt werden. Brodnig erläutert, dass Trolle sich derselben Taktik bedienen, indem sie verletzende Äußerungen verbreiten und hoffen, dass jemand anbeißt. Trolle sind dafür bekannt, sich in Diskussionsforen irreführend, destruktiv und störend zu verhalten (Buckels, Trapnell, & Paulhus, 2014). Sie fühlen sich gerne intellektuell überlegen, provozieren mit naiven und dummen Fragen und erfreuen sich am Leid anderer (Brodnig, 2016). Außerdem versuchen sie bewusst, Diskussionen im Internet zu stören und Konflikte zu provozieren bzw. zu verschärfen (Hardaker, 2010). Zu den Strategien der Trolle zählen laut der Linguistin Claire Hardaker: (1) das absichtliche Abschweifen vom Thema, (2) überzogene Kritik, beispielsweise in Form von Hinweisen auf Tippfehler, (3) das Einnehmen einer Gegenposition, um Andersdenkende bewusst zu provozieren, (4) das Gefährden anderer durch gefährliche Ratschläge bzw. das Vortäuschen einer Gefahr, wodurch andere gezwungen werden zu reagieren, um einen Schaden abzuwenden, (5) Schocken, u.a. durch unsensible und unangemessene Äußerungen zu sensiblen Themen oder explizite und anstößige Äußerungen zu Tabuthemen und (6) grundlose Angriffe auf andere in Form von Beleidigungen oder Drohungen (Hardaker, 2013).

Neben den Trollen identifiziert Ingrid Brodnig in ihrem Buch auch einen zweiten Typus problematischer Internetnutzer. Die sogenannten „Glaubenskrieger“ sind in Diskussionsforen oft respektlos und verhalten sich gegenüber Andersdenkenden äußerst aggressiv, weil sie von einer Idee dermaßen überzeugt sind, dass sie gegenteilige Meinungen von anderen Diskussionsteilnehmern nicht dulden. In Diskussionsforen suchen Glaubenskrieger die Konfrontation. Sie versuchen andere Diskussionsteilnehmer von der einzig wahren „Wahrheit“ zu überzeugen, weil sie sich für besser informiert halten. Sie möchten die naive und befangene Gesellschaft instruieren und sind gegen alle noch so stichhaltigen Gegenargumente immun. Personen, die gegenteiliger Meinung sind, werden oft als Verblendete und Lügner bezeichnet, obwohl Glaubenskrieger selber Lügen



einsetzen und sich auf unseriöse Quellen berufen, um ihre Standpunkte zu untermauern. Glaubenskrieger fallen letztendlich auch durch ihre aggressive Tonalität auf. Durch ständige Beleidigungen und Provokationen wollen sie erreichen, dass sich Andersdenkende nicht mehr zu Wort melden und aus der Diskussion zurückziehen (Brodnig, 2016).

Aus psychologischer Sicht lassen sich laut der Medienpsychologin Josephine B. Schmitt (2017) unterschiedliche Beweggründe für die Verbreitung von Hass im Netz unterscheiden. Häufig werden Personen und Personengruppen, die ein gemeinsames Merkmal (Herkunft, Religion, sexuelle Orientierung etc.) aufweisen, gezielt mit Beleidigungen und Herabwürdigungen ausgegrenzt. Durch die Abwertung der Fremdgruppe soll die eigene Gruppenidentität gestärkt werden. Eine Erklärung für dieses Verhalten liefert der Soziologe William Graham Sumner. In seinem Buch „Folkways“ erklärt er, dass die eigene Gruppe oft idealisiert wird, während die Fremdgruppe abgewertet und als minderwertig angesehen wird. Sumner nennt dieses Konzept „Ethnozentrismus“ und definiert es wie folgt: „Ethnocentrism is the technical name for this view of things in which one's own group is the center of everything, and all others are scaled and rated with reference to it.“ (Sumner, 2007, S. 13). Es ist durchaus naheliegend, dass eine Abgrenzung bzw. Abwertung einer Fremdgruppe schnell zu vorurteilsgeleitetem Hass führen kann.

Ein weiterer Grund für die Verbreitung von Hass im Internet ist Wut, die sich laut Schmitt (2017) oft auf ein Gefühl der Bedrohung zurückführen lässt. Zwar ist die Bedrohung oft unbegründet, dennoch scheinen beispielsweise unbekannte Situationen derartige Gefühle auszulösen. Der Sozialpsychologe Franz Asbrock erklärt in einem Interview mit dem *Tagesspiegel*, dass die Flüchtlingszuwanderung 2015 bei vielen Menschen derartige Bedrohungsgefühle ausgelöst hat. Ihren Unmut darüber haben sie dann im Internet zum Ausdruck gebracht. Laut Asbrock ist das eine übliche Form, wie mit Bedrohung und Angst umgegangen wird (Tagesspiegel.de, 2018). Schmitt ist ebenfalls der Meinung, dass das Gefühl der Bedrohung bei diesen Personen erst abnimmt, wenn sie ihre Wut bzw. ihren Hass ausgedrückt und die für die (vermeintliche) Bedrohung zuständige Fremdgruppe eingeschüchtert haben.

Josephine B. Schmitt (2017) erklärt außerdem, dass viele Personen Hass im Netz verbreiten, um Dominanz und Macht in gesellschaftlichen Diskursen zu demonstrieren. Ihr Ziel ist es, andere Diskussionsteilnehmerinnen und Diskussionsteilnehmer von ihrer

Meinung zu überzeugen und letztendlich dazu zu verleiten, ebenfalls Hasskommentare zu verbreiten.

Als ein weiteres Motiv für Hassrednerinnen und Hassredner nennt Schmitt die Freude am Beleidigen und Erniedrigen anderer. Ingrid Brodnig erklärt in ihrem Buch, dass gehässige Internetuserinnen und Internetuser andere Personen aus purer Belustigung beleidigen. Diese Schadenfreude wird von den Trollen „LULZ“ („I did it for the LULZ“) genannt. „LULZ“ ist eine Abwandlung von „LOL“ („laughing out loud“) und soll das grundlos provozierende Verhalten der Trolle erklären (Brodnig, 2016). Eine wissenschaftliche Erkenntnis für das Verhalten von Trollen konnten kanadische Wissenschaftlerinnen und Wissenschaftler liefern. Die 2014 durchgeführte Studie „Trolls just want to have fun“ (Buckels, Trapnell, & Paulhus) hat sich mit den Persönlichkeitsmerkmalen eines Internet-Trolls beschäftigt. Erin Buckels und ihre Kollegen haben dabei konkret den Einfluss von verschiedenen Persönlichkeitsmerkmalen auf das Kommentierungsverhalten im Internet analysiert. Sie haben untersucht, inwiefern Trolle folgende negative Persönlichkeitsmerkmale der Dunklen Tetrad aufweisen: Sadismus, Narzissmus, Psychopathie und Machiavellismus. Sie konnten nachweisen, dass Trolle besonders häufig sadistische Charakterzüge aufweisen, weil sie es genießen, andere Personen zu provozieren.

## 2.3. Faktoren, die Hass im Netz begünstigen

Wissenschaftlerinnen und Wissenschaftler interessieren sich neben den Beweggründen für die Verbreitung von Hass im Netz auch für die unterschiedlichen Faktoren, die Hass im Netz begünstigen. Sie suchen nach Antworten auf die Frage, weshalb gerade online so viel Hass verbreitet wird. Ein wesentlicher Grund für das vermehrte Auftreten von Hass im Netz ist die Anonymität und das Gefühl, sich in einem rechtsfreien Raum zu bewegen. Es werden allerdings auch einige internetspezifische Faktoren genannt, die die Verbreitung von Hass im Netz fördern und die im Folgenden näher erläutert werden sollen.

### 2.3.1. Anonymität

Die unendlichen Weiten des Internets bergen leider auch ihre Schattenseiten. Unter dem scheinbaren Deckmantel der Anonymität lassen viele Menschen ihren Unmut im Netz aus. Wieso dieses Phänomen gerade im Internet so präsent ist, hat mehrere Gründe. John Suler (2004) nennt die Enthemmung im Internet als einen davon. Folgende Faktoren spielen dabei laut seiner „Online Disinhibition Effect“-Theorie eine besonders große Rolle:

- Die Anonymität: ist einer der Hauptfaktoren für den Online-Enthemmungseffekt. Wenn die Identität nicht preisgegeben werden muss, fühlt man sich nicht so verwundbar.
- Die Unsichtbarkeit: ist aufgrund des fehlenden Augenkontakts gegeben. Dadurch fällt es den Menschen leichter, ihren Unmut zu äußern.
- Die Asynchronität: Menschen interagieren nicht in Echtzeit. Eine Reaktion auf eine beleidigende Äußerung ist daher auch nicht gleich zu erwarten.
- Die solipsistische Introjektion: Durch den fehlenden Augenkontakt wird das Gegenüber nicht vollständig wahrgenommen. Daraufhin kreiert man eine Art virtuellen Charakter und vervollständigt das Gegenüber in der eigenen Phantasie. Und in der Phantasie tut und sagt man Dinge, die man im echten Leben nicht tun und sagen würde.

- Die dissoziative Vorstellungskraft: Die Vermischung von Realem und Fiktion führt zu dem Glauben, dass alles nur ein Spiel ist und die Regeln der Kommunikation, wie man sie aus dem Alltag kennt, hier keine Anwendung finden.
- Das Fehlen von Autorität: begünstigt enthemmte Kommentare.

Lapidot-Lefler & Barak (2012) haben ebenfalls herausgefunden, dass die Hemmschwelle im Internet vor allem auch durch den fehlenden Augenkontakt sehr gering ist. Sie konnten nachweisen, dass bei Blickkontakt viel seltener beleidigt und gedroht wird und behaupten deshalb, dass fehlender Blickkontakt zu einem Gefühl der sogenannten „Unidentifizierbarkeit“ führt, welche wiederum die Hemmschwelle senkt.

Ein Jahr später (2013) haben sich Cho & Acquisti gefragt, ob der Grad der Identifizierbarkeit einen Einfluss auf das Kommentierungsverhalten auf sozialen Netzwerken hat. Die Wissenschaftler haben herausgefunden, dass beleidigende Äußerungen seltener vorkommen, wenn Benutzer und Benutzerinnen ihre Identität preisgeben müssen. Das bestätigt die Annahme, dass Anonymität eine wesentliche Rolle spielt. Wer sich anonym und unsichtbar fühlt, postet auch enthemmter.

### 2.3.2. Rechtsfreier Raum

Ein weiterer Faktor, der Hass im Netz begünstigt, ist der Glaube der Hassposterinnen und Hassposter, sich in einem rechtsfreien Raum zu bewegen. Zwar gibt es durchaus unterschiedliche Möglichkeiten, sich gegen Hass im Netz zu wehren, allerdings wissen viele Betroffene oft nicht über ihre Rechte Bescheid. Hier mangelt es leider immer noch an Aufklärung und Sensibilisierung. Das Internet wird oft auch als rechtsfreier Raum wahrgenommen, weil es sehr schwer ist, die Täterinnen und Täter ausfindig zu machen, vor allem wenn sie ihre Hassbotschaften anonym verbreiten (Windhager, 2016). Doch der Schein trügt. Zwar ist die Rechtsdurchsetzung bei anonymen Hassposterinnen und Hasspostern tatsächlich erschwert, dennoch ist sie nicht unmöglich. Nach dem österreichischen E-Commerce-Gesetz (§ 18 Abs 4 ECG) sind nämlich beispielsweise Betreiber von Online-Diskussionsforen, die im Gesetz als Host-Provider bezeichnet werden, auf Verlangen Dritter zur Herausgabe von Userdaten verpflichtet, „sofern diese ein überwiegendes rechtliches Interesse an der Feststellung der Identität eines Nutzers und eines bestimmten rechtswidrigen Sachverhalts sowie überdies glaubhaft machen,

dass die Kenntnis dieser Informationen eine wesentliche Voraussetzung für die Rechtsverfolgung bildet.“ (Rechtsinformationssystem Bundeskanzleramt Österreich, 2019a). Auskunftspflichtig sind allerdings nur Host-Provider, nicht jedoch Betreiber von Suchmaschinen, Linksetzer oder Domainverwaltungsstellen (Zankl, 2016).

In Österreich müssen Host-Provider wie Facebook bei Rechtsverstößen Auskunft über Namen, Adresse ggf. auch E-Mail-Adresse (Zankl, 2016), nicht jedoch über die IP-Adresse geben (Windhager, 2016). In Frankreich hingegen will Facebook erstmals IP-Adressen von Userinnen und Usern, die Hassbotschaften verbreiten, an französische Gerichte weiterleiten. Bisher wurden Benutzerdaten nur zur Terrorismusbekämpfung weitergegeben, künftig will Facebook jedoch enger mit der Justiz zusammenarbeiten, um verstärkt gegen Hass im Netz vorzugehen (derStandard.at, 2019a). Je mehr auf nationaler und internationaler Ebene (durch Gesetze, Kooperationen mit Unternehmen etc.) getan wird, umso wahrscheinlicher ist es, dass das Internet nicht mehr als rechtsfreier Raum wahrgenommen wird.

### 2.3.3. Nasty-Effect

Eine weitere interessante Erkenntnis lieferten Kommunikationswissenschaftlerinnen und Kommunikationswissenschaftler der University of Wisconsin-Madison (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014). Sie gingen der Frage nach, inwiefern grobe und unhöfliche Kommentare einen Einfluss auf passive Teilnehmerinnen und Teilnehmer bzw. Leserinnen und Leser einer Online-Diskussion haben. Herausgefunden werden sollte auch, ob solche Kommentare beeinflussen können, wie Leserinnen und Leser über das diskutierte Thema denken. Hierfür haben die Forscherinnen und Forscher 1183 Probandinnen und Probanden einen neutral verfassten Artikel einer kanadischen Zeitung zum Thema Nanotechnologie lesen lassen, inklusive der darunter verfassten Kommentare von anderen Leserinnen und Lesern. Diese Kommentare wurden unterschiedlich manipuliert und in höfliche und unhöfliche Kommentare unterteilt. Die Argumente waren in beiden Gruppen gleich kritisch, die unhöflichen Kommentare beinhalteten lediglich zusätzliche Beleidigungen. Anschließend wurden die Befragten (wie bereits zuvor) zu ihrer Einstellung zum Thema Nanotechnologie befragt. Die Beleidigungen hatten tatsächlich einen erheblichen Einfluss auf die Meinung der Befragten. Die Beschimpfungen haben zu einer Polarisierung geführt, da sich die Meinungen jener Befragten aus der Gruppe mit

den unhöflichen Kommentaren verstärkt haben. Befürworterinnen und Befürworter waren überzeugter als zuvor, Gegnerinnen und Gegner deutlich abgeneigter. Die Gruppe mit den höflichen Kommentaren war hingegen weniger gespalten. Diese Studie hat somit gezeigt, dass Leserinnen und Leser von Beleidigungen beeinflusst werden, indem sie sie in ihrer eigenen Meinung bestärken. Durch diesen Effekt, den die Kommunikationswissenschaftlerinnen und Kommunikationswissenschaftler den „Nasty Effect“ nennen, wird das Diskussionsklima verschlechtert und eine sachliche Argumentation unmöglich.

#### 2.3.4. Echokammern und Filterblasen

Ein weiteres Problem des Internets ist, dass es die Entstehung von sogenannten Echokammern fördert. Laut Ingrid Brodnig sind Echokammern „digitale Räume, in denen Nutzer hauptsächlich Inhalte eingeblendet bekommen, die ihre Meinung bekräftigen.“ (Brodnig, 2016, S. 22). Das bedeutet, dass Personen, die sich in diesen Echokammern befinden, nie mit Andersdenkenden konfrontiert werden und somit nie die Möglichkeit haben, ihre Einstellungen, Gedanken und Werte kritisch zu hinterfragen. Echokammern können somit Hass im Netz begünstigen, weil Hassposterinnen und Hassposter dadurch in ihrem Hass bestärkt werden. Ihr Verhalten wird nämlich zu keinem Zeitpunkt infrage gestellt bzw. kommen sie nie mit Informationen und Argumenten in Kontakt, die eine Veränderung in ihrem Verhalten bewirken könnten.

Die Filterblase funktioniert ähnlich wie eine Echokammer. Eine Echokammer entsteht, wenn Personen eigenständig entscheiden, welche Informationen und (Medien)inhalte sie konsumieren wollen. Die Filterblase hingegen entsteht durch eine Informationsselektion von Algorithmen (Brodnig, 2016). Eli Pariser erklärt in seinem Buch „The Filter Bubble: What the Internet is Hiding from You“ (2011) wie diese Algorithmen funktionieren:

„The new generation of Internet filters looks at the things you seem to like – the actual things you’ve done, or the things people like you like – and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you’ll do and want next. Together, these engines create a unique universe of information for each of us – what I’ve come to call a filter bubble – which fundamentally alters the way we encounter ideas and information.“ (Pariser, 2011, S. 9)

Wenn Userinnen und User somit beispielsweise auf Facebook eine bestimmte Seite liken, zieht der Algorithmus daraus seine Schlüsse. Er richtet sich nach den Interessen der Userinnen und User und blendet dementsprechend relevante Inhalte ein und irrelevante Inhalte aus (Brodnig, 2016).

Eines hat die Filterblase mit den Echokammern gemeinsam. Beide Effekte können Hass im Netz begünstigen, weil durch sie personalisierte Informationen und Inhalte an Personen mit dem gleichen Weltbild verbreitet werden. Bei diesen Personen entsteht dann der Eindruck, dass die überwiegende Mehrheit ihre Gedanken, Ansichten und Einstellungen teilt. Gerade bei Falschmeldungen kann das jedoch schwerwiegende Folgen haben, wie folgendes Beispiel zeigt. 2015 kam erstmals das Gerücht auf, dass Asylwerberinnen und Asylwerber kostenlos Smartphones von Hilfsorganisationen wie der Caritas erhalten würden. Wie *derStandard.at* berichtet, wurde auf der Facebook-Seite „Wir unterstützen Norbert Hofer“ folgende Behauptung gepostet: „Asylant kauft auf Kosten der Caritas ein I-Phone um 900,- ! Anruf bei CARITAS. KEIN PROBLEM! NIE MEHR SPENDEN FÜR DIESEN VERRÄTERVEREIN!!! STOPP DER ASYLINDUSTRIE!!!“ (*derStandard.at*, 2016a). Diese Fehlmeldung wurde zunächst aufgrund des Algorithmus jenen Personen angezeigt, die zuvor bereits nach ähnlichen Informationen gesucht hatten oder denen, die die Facebook-Seite, auf der das Posting geteilt wurde, gefällt. Viele dieser Personen haben anschließend das Posting auf ihrer eigenen Seite geteilt. Laut *derStandard.at* wurde das Posting vor der Löschung von der ursprünglichen Facebook-Seite insgesamt über 1.600 Mal von anderen Userinnen und Usern geteilt. Anhand dieses Beispiels wird auch deutlich, weshalb die Verbreitung von Fehlmeldungen derart problematisch ist. Personen, die das besagte Posting auf Facebook gelesen haben, werden auch weiterhin nur Postings dieser Art zu lesen bekommen. Sie werden dementsprechend nie mit Postings, die das Gegenteil behaupten bzw. belegen können, in Berührung kommen. Das Internet, das die Entstehung von Echokammern und Filterblasen fördert, kann somit Hass im Netz wesentlich begünstigen. Vor allem bei polarisierenden Themen können dabei vermehrt aggressive und hasserfüllte Kommentare hervorgerufen werden.

## 2.4. Auswirkungen von Harmful Online Communication

Wie bereits erwähnt, beschreibt „Harmful Online Communication“ das Phänomen „Hass im Netz“ deshalb sehr treffend, weil der Begriff die Auswirkungen der aggressiven und hasserfüllten Sprache auf die Betroffenen beinhaltet. Doch inwiefern ist Online-Kommunikation schädlich? Welche Auswirkungen kann HOC auf die Betroffenen haben? Bisher haben sich die Studien, die zum Thema Hass im Netz durchgeführt wurden, vergleichsweise eher wenig mit den Auswirkungen von Online Hate Speech bzw. Harmful Online Communication auseinandergesetzt. Fest steht allerdings, dass aggressive und hasserfüllte Äußerungen dazu führen können, dass den Betroffenen emotional die Menschlichkeit aberkannt wird (Baldauf, Banaszczuk, Koreng, Schramm, & Stefanowitsch, 2015). Denn wenn Personen aufgrund ihrer Herkunft, Religion oder sexuellen Orientierung verunglimpft werden, werden sie nicht mit ausreichender Würde und Respekt behandelt. Dass Auswirkungen bei Personen, die im Netz konstant hasserfüllten Äußerungen ausgesetzt sind, weit über das genannte hinausgehen, erklärt Diplompsychologin Dorothee Scholz in einem Interview mit der Amadeu-Antonio-Stiftung:

„Die möglichen Auswirkungen eines solchen Dauerbeschusses reichen von Gefühlen der Hilflosigkeit, Angst, Scham, starken Verunsicherung und generell emotionalen Belastung über sozialen Rückzug und körperliche Erkrankungen bis hin zu psychischen Störungen und sogar Selbsttötung. Anhaltende Bedrohungen dieser Art können außerdem die Persönlichkeit verändern, lebenslange Verbitterung hervorrufen oder jemanden dazu bringen, sich emotional über Suchtverhalten zu schützen. Die Verletzungen sind so gravierend, dass viele Menschen in Befragungen sogar angeben, bereitwilliger körperliche als psychische Gewalt ertragen zu wollen.“ (Baldauf, Banaszczuk, Koreng, Schramm, & Stefanowitsch, 2015, S. 26)

Die Auswirkungen können von häufig auftretenden Gefühlen der Ohnmacht bis hin zu gravierenden psychischen Störungen reichen. Psychische Störungen können sich dann entwickeln, wenn Betroffene die erlebten Beleidigungen und verbalen Angriffe nicht verarbeiten können (Baldauf, Banaszczuk, Koreng, Schramm, & Stefanowitsch, 2015).

Die australische Studie von Gelber & McNamara (2016) hat ebenfalls bewiesen, dass hasserfüllte Äußerungen unterschiedliche psychische und physische Folgen für die Betroffenen haben können. Aus den 101 durchgeführten qualitativen Interviews mit



Indigenen und ethnischen Minderheiten geht hervor, dass die Betroffenen häufig unter einer Art existenziellem Schmerz leiden und Gefühle von Angst, Entmachtung, Ausgrenzung, Entmenschlichung, Wut und Frustration verspüren.

Nun stellt sich die Frage, ob Harmful Online Communication auch über die genannten psychischen und physischen Folgen der Betroffenen hinausreichen kann. Kann digitaler Hass beispielsweise auch zu realer Gewalt führen? Laut der Psychologin Dorothee Scholz ist bei Hassposterinnen und Hasspostern, die online anderen Personen Gewalt androhen, die psychische Hemmschwelle zur Gewaltausübung gesenkt (Baldauf, Banaszczuk, Koreng, Schramm, & Stefanowitsch, 2015). Aber führt eine niedrige Hemmschwelle tatsächlich zu gewalttätigen Zwischenfällen? Zwei Wissenschaftler der University of Warwick (Müller & Schwarz, 2018) haben in ihrer Studie den Zusammenhang zwischen Online Hate Speech auf Facebook und Hassverbrechen in Form von realer Gewalt gegen Geflüchtete untersucht. Müller & Schwarz behaupten, nachgewiesen zu haben, dass das soziale Netzwerk für reale, gewalttätige Zwischenfälle verantwortlich ist. Denn in den Gebieten Deutschlands, wo viele hasserfüllte Kommentare gegen Geflüchtete gepostet wurden, sei auch die Anzahl realer Gewalt auf Geflüchtete etwa in Form von Brandstiftung oder Körperverletzung höher. Doch sind die Ergebnisse der Studie von Müller & Schwarz mit Vorsicht zu genießen. Die Studie, die vor der Veröffentlichung keinem Peer-Review-Verfahren unterzogen wurde, stößt auf Kritik. Jonas Kaiser, der an der Harvard University über digitale und politische Kommunikation forscht, sagt beispielsweise, „dass ein Zusammenhang hergestellt wird, der durch die Studie nicht gegeben wird“ (SZ.de, 2018). Die Forscher haben falsche Schlüsse gezogen, indem sie behauptet haben, dass eine Kausalität zwischen hasserfüllten Kommentaren auf Facebook und gewalttätigen Zwischenfällen besteht. Tatsächlich konnten sie aber nur eine Korrelation nachweisen, die allerdings nichts über die Kausalität aussagt. Ein weiterer Kritikpunkt besteht laut Kaiser darin, dass die gezogene Stichprobe der Forscher nicht repräsentativ bzw. aussagekräftig war. Die Aktivitäten der Facebookseite der AfD (der rechtspopulistischen Partei „Alternative für Deutschland“) wurden nämlich mit der Facebook-Seite von Nutella verglichen. Bei der Seite von Nutella handelte es sich allerdings um die internationale Facebookseite mit 32 Millionen Fans, von denen letztendlich nur 21.915 aus Deutschland stammen (SZ.de, 2018).

Dass sich Hass im Netz schnell verbreitet und dadurch Personen sowohl in ihren Gedanken, Ansichten und Einstellungen als auch in ihrer Wut und ihrem Ärger bestärkt, wurde bereits mit dem Filterblasen-Effekt erklärt (siehe Kapitel 2.3.4.). Für die Annahme,

dass Hass im Netz auch physische Gewalt zur Folge haben kann, gibt es hingegen bisher keine wissenschaftlichen Belege.

## 2.5. Gesetzliche Grundlagen

Da erwiesenermaßen HOC schwerwiegende Folgen auf die Betroffenen haben kann, stellt sich die Frage, wie rechtlich dagegen vorgegangen werden kann bzw. welche gesetzlichen Vorschriften es bereits gibt, die HOC in Österreich, Deutschland, den Vereinigten Staaten und Großbritannien einschränken. Interessant ist dabei vor allem, wo in den verschiedenen Ländern die freie Meinungsäußerung aufhört und wo andere strafrechtlich relevante Tatbestände anfangen.

### 2.5.1. Das Recht auf freie Meinungsäußerung

Eine der wichtigsten Grundlagen in einer Demokratie ist das Recht auf freie Meinungsäußerung. Das die Meinungsfreiheit allerdings nicht grenzenlos ist, geht unter anderem aus der Europäischen Menschenrechtskonvention (EMRK) hervor. Diese enthält einen Katalog von Grundrechten und Menschenrechten, an den sich alle 47 Mitgliedstaaten des Europarates, darunter auch Österreich, Deutschland und das Vereinigte Königreich, halten müssen (Europarat, 2019). Freie Meinungsäußerung wird laut Art. 10 Abs. 1 der Europäischen Menschenrechtskonvention wie folgt definiert:

„Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.” (European Convention on Human Rights, 1950, S. 12)

Geschützt wird demnach durch Art. 10 Abs. 1 sowohl die Freiheit der Meinungsäußerung und -bildung als auch die Informationsfreiheit und die Medienfreiheit. Das Recht auf freie Meinungsäußerung ist allerdings nicht grenzenlos. Schranken ergeben sich aus dem materiellen Gesetzesvorbehalt des Art. 10 Abs. 2, welcher folgendes besagt:

„The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of

national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.” (European Convention on Human Rights, 1950, S. 12)

Eine Einschränkung des Rechts auf freie Meinungsäußerung ist demnach laut Art. 10 Abs. 2 EMRK bei wichtigen öffentlichen (z.B. nationale und öffentliche Sicherheit, Aufrechthaltung der Ordnung und der Verbrechensverhütung) und privaten (z.B. Schutz des guten Rufes oder der Rechte anderer) Interessen gerechtfertigt. Eingriffe sind allerdings nur dann zulässig, wenn sie gesetzlich vorgesehen sind, einem oder mehreren Zwecken aus Abs. 2 dienen und verhältnismäßig sind (Soll, 2001).

Für die vorliegende Arbeit ist besonders interessant, wo und wie die Grenze zwischen freier Meinungsäußerung und HOC gezogen wird und ob es länderspezifische Unterschiede im Umgang mit HOC gibt. Grundsätzlich gilt, dass die Meinungsfreiheit nur so weit reicht, bis die Rechte anderer verletzt werden (Brodnig, 2016). Ob eine Äußerung noch zulässig ist oder bereits rechtswidrig ist, entscheiden in Europa zunächst die nationalen Gerichte und in letzter Instanz auch der Europäische Gerichtshof für Menschenrechte (EGMR). Der EGMR, der über Verletzungen der Europäischen Menschenrechtskonvention urteilt, zieht die Grenze zur Meinungsäußerungsfreiheit wie folgt: Aussagen, die nur beleidigen, schockieren oder stören, stellen keine Verletzung der Meinungsäußerungsfreiheit dar (EGMR, 1976). Handlungen, die hingegen durch die EMRK gewährte Rechte missbrauchen, um Grund- und Menschenrechte von Personen oder bestimmten Personengruppen zu verletzen, werden laut Art. 17 EMRK nicht geschützt (European Convention on Human Rights, 1950). In den Vereinigten Staaten ist für die Beurteilung, ob eine Äußerung in den Schutzbereich der freien Meinungsäußerung fällt, die Rechtsprechung des Supreme Court entscheidend.

Neben den rechtlichen Institutionen interessiert sich auch die Wissenschaft für diese Thematik. In diversen Publikationen gehen Wissenschaftlerinnen und Wissenschaftler der Frage nach, wo bzw. ob eine Grenze zwischen freier Meinungsäußerung und strafbaren Äußerungen gezogen werden sollte. Dabei halten viele zunächst fest, dass es sich bei der Meinungsäußerungsfreiheit um eine unerlässliche Grundfreiheit handelt, die für die Aufrechthaltung der Demokratie von zentraler Bedeutung ist (Tsesis, 2009). In einer liberalen Demokratie sollten Personen uneingeschränkt die Möglichkeit haben, sich über das politische Geschehen zu unterhalten und Kritik auszuüben, ohne Konsequenzen

befürchten zu müssen. Nichtsdestotrotz sind die Autoren überwiegend der Meinung, dass hasserfüllte Äußerungen nicht von der Meinungsfreiheit geschützt werden sollten. Ein Argument hierfür lautet, dass Hassreden die Grundwerte einer liberalen demokratischen Gesellschaft verletzen, weil sie inhaltlich nicht so bedeutsam sind wie andere Inhalte, die zurecht von der Meinungsäußerungsfreiheit geschützt werden (Sirsch, 2013). Ein weiterer Grund, der für eine strenge Regulierung von HOC spricht, ist die wissenschaftlich erwiesene schädliche Wirkung auf die Betroffenen (Waldron, 2010; Sirsch, 2013). Durch die beleidigenden Äußerungen werden die Betroffenen in ihren Persönlichkeitsrechten verletzt. Gerade weil in einer liberalen demokratischen Gesellschaft Persönlichkeitsrechte genauso geschützt werden wie die freie Meinungsäußerung, ist im Zweifelsfall zumindest in Europa eine Interessensabwägung durch ein Gericht durchzuführen.

Wie die nationalen (und internationalen) Gerichte im Zweifelsfall entscheiden, hängt stark von den jeweiligen Gesetzen ab. Grundsätzlich kommt der Meinungsäußerungsfreiheit in vielen demokratischen Ländern ein besonders hoher Stellenwert zu. Einschränkungen der Meinungsäußerungsfreiheit sind nur zum Schutz öffentlicher oder privater Interessen erlaubt. Im internationalen Vergleich können jedoch Unterschiede in der Einschränkung der Meinungsäußerungsfreiheit festgestellt werden (Einwiller & Kim, 2018). Um herauszufinden, wie die freie Meinungsäußerung in unterschiedlichen Ländern bewertet wird, hat das Pew Research Center (2015a) einen Index für freie Meinungsäußerung („Free Expression Index“) entwickelt. Hierfür wurden Personen aus 38 Ländern mittels persönlichen und Telefoninterviews zu ihren Einstellungen zu unterschiedlichen Aspekten der Meinungsfreiheit befragt. Laut der Studie hatte die Meinungsäußerungsfreiheit für die Befragten aus den Vereinigten Staaten den höchsten und für die Befragten aus dem Senegal den geringsten Stellenwert. Für die vorliegende Forschungsarbeit sind die Werte aus Deutschland, den Vereinigten Staaten und Großbritannien relevant. Wie aus Abbildung 1 ersichtlich ist, sprechen sich die befragten Personen aus Deutschland noch am ehesten (4,34) für die Einschränkung der Meinungsäußerungsfreiheit unter bestimmten Umständen aus. In den Vereinigten Staaten sind die Befragten hingegen eher (5,73) der Meinung, dass die Meinungsäußerungsfreiheit auch dann nicht eingeschränkt werden sollte, wenn Minderheiten beleidigt werden oder Personen zu gewalttätigen Protesten aufrufen. Das Vereinigte Königreich (4,78) tendiert im Vergleich zu Deutschland eher dazu, der freien Meinungsäußerung den höheren Stellenwert beizumessen. In dieser Studie wurden keine Einstellungen zur freien Meinungsäußerung von Personen aus Österreich erhoben. Es kann allerdings angenommen werden, dass die Werte ähnlich wie in Deutschland ausgefallen wären.

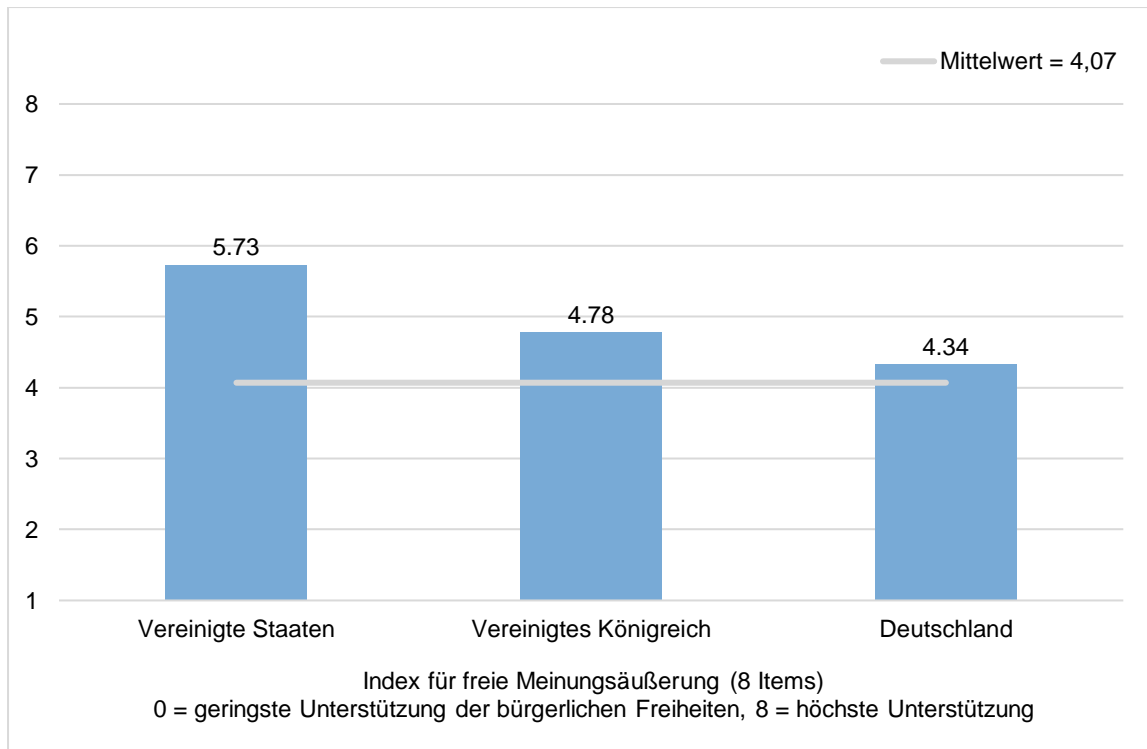


Abbildung 1: Einstellungen zur freien Meinungsäußerung (Auswahl), Quelle: Pew Research Center, „Global Attitudes Survey“ (2015a)

Eine weitere Untersuchung des Pew Research Center (2015b) hat ergeben, dass Personen aus europäischen Ländern (darunter auch Deutschland und dem Vereinigten Königreich) mit 49% eher bereit sind, Beschränkungen von beleidigenden bzw. anstößigen Äußerungen hinzunehmen als Personen aus den Vereinigten Staaten (28%). Aus Abbildung 2 geht hervor, wie sich bei dieser Frage die Meinungen der Länder scheiden. In Deutschland ist die überwiegende Mehrheit (70%) der Meinung, dass anstößige Äußerungen verboten und dementsprechend nicht durch die Meinungsäußerungsfreiheit geschützt werden sollten. Nur 27% sind hingegen der Ansicht, dass anstößige Aussagen über Minderheiten erlaubt sein sollten. Personen aus dem Vereinigten Königreich tendieren hingegen eher (54%) dazu, anstößige Aussagen gegenüber Minderheiten zu tolerieren. An den Ergebnissen der Vereinigten Staaten ist wieder deutlich zu erkennen, dass die überwiegende Mehrheit der befragten Personen (67%) der freien Meinungsäußerung einen höheren Stellenwert beimisst als dem Verbot von anstößigen Äußerungen (28%). Auch bei dieser Untersuchung wurden Einstellungen aus Österreich nicht berücksichtigt.

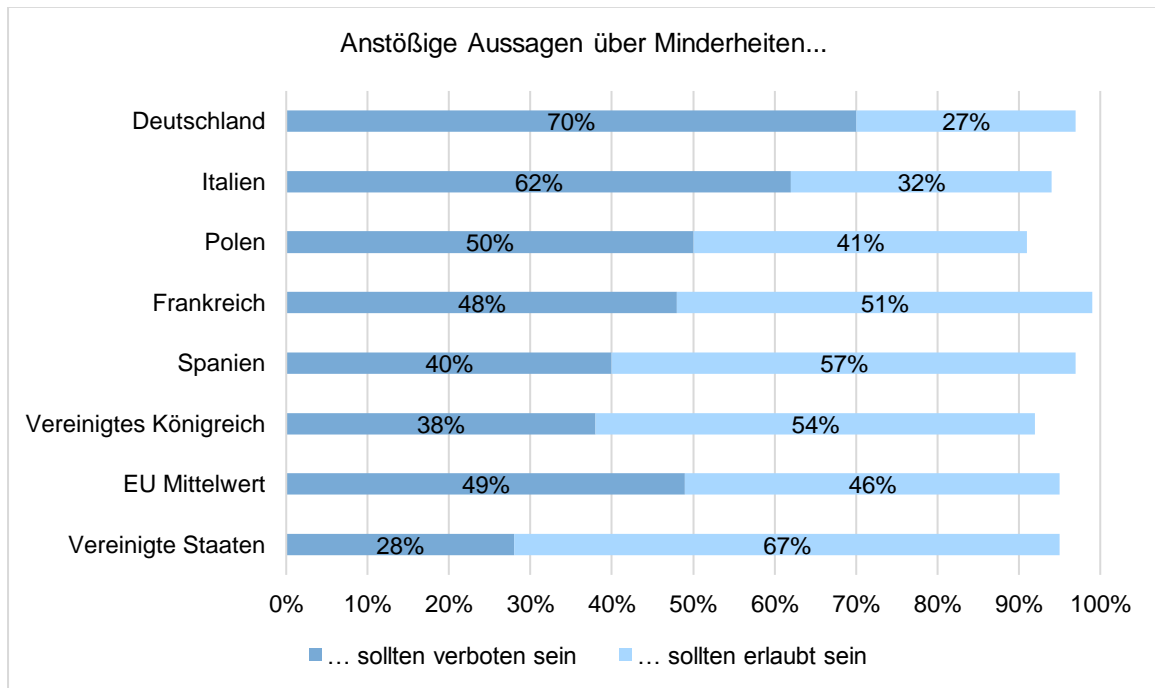


Abbildung 2: Einstellungen zur Zensur von Aussagen, die gegenüber Minderheiten anstößig sind, Quelle: Pew Research Center, „Global Attitudes Survey“ (2015b)

Ob sich die Ergebnisse des Pew Research Center in den Gesetzen der für die vorliegende Forschungsarbeit relevanten Ländern widerspiegeln und wo in Österreich, Deutschland, den Vereinigten Staaten und Großbritannien die Grenze zwischen freier Meinungsäußerung und HOC gezogen wird, soll in den folgenden Kapiteln erläutert werden.

## 2.5.2. Gesetzliche Lage in Österreich

In Österreich ist die freie Meinungsäußerung zunächst im Staatsgrundgesetz (StGG) geregelt. Laut Art. 13 StGG hat jedermann „das Recht, durch Wort, Schrift, Druck oder durch bildliche Darstellung seine Meinung innerhalb der gesetzlichen Schranken frei zu äußern“ (Rechtsinformationssystem Bundeskanzleramt Österreich, 2019b). Außerdem ist das Recht auf freie Meinungsäußerung in Art. 10 der Europäischen Menschenrechtskonvention, die in Österreich Verfassungsrang hat und unmittelbar anwendbar ist, verankert.

In Österreich wird die Grenze zwischen freier Meinungsäußerung und HOC dort gezogen, wo Rechte anderer verletzt werden (Brodnig, 2016). Äußerungen, die nicht von der Meinungsäußerungsfreiheit geschützt werden, sind vor allem im Strafgesetzbuch (StGB) zu finden. Zu den strafbaren Delikten zählen unter anderem die Drohung, Beleidigung oder Verhetzung.

Von einer „gefährlichen Drohung“ spricht man laut § 107 StGB, wenn eine Person durch eine Drohung in „Furcht und Unruhe“ versetzt wird (Rechtsinformationssystem Bundeskanzleramt Österreich, 2015a). Laut der Rechtsanwältin Maria Windhager ist bei den meisten Drohungen im Internet der Tatbestand der gefährlichen Drohung nicht erfüllt, weshalb die drohenden Personen nicht nach § 107 StGB verurteilt werden können (Windhager, 2016). Um derartige Drohungen dennoch bestrafen zu können, wurde 2005 zusätzlich die „Fortgesetzte Belästigung im Wege einer Telekommunikation oder eines Computersystems“, also Cyber-Mobbing, in das StGB aufgenommen. § 107c StGB stellt unzumutbare und über einen längeren Zeitraum hindurch fortgesetzte Ehrverletzungen im Internet und das Verbreiten von Bildaufnahmen des höchstpersönlichen Lebensbereiches einer Person ohne deren Zustimmung unter Strafe (Rechtsinformationssystem Bundeskanzleramt Österreich, 2019c).

Ein weiterer strafrechtlich relevanter Tatbestand ist die „üble Nachrede“. Wer laut § 111 StGB „einen anderen in einer für einen Dritten wahrnehmbaren Weise einer verächtlichen Eigenschaft oder Gesinnung zeiht oder eines unehrenhaften Verhaltens oder eines gegen die guten Sitten verstößenden Verhaltens beschuldigt, das geeignet ist, ihn in der öffentlichen Meinung verächtlich zu machen oder herabzusetzen, ist mit Freiheitsstrafe bis zu sechs Monaten oder mit Geldstrafe bis zu 360 Tagessätzen zu bestrafen.“ (Rechtsinformationssystem Bundeskanzleramt Österreich, 2018a). Laut § 111 Abs. 2 StGB erhöht sich das Strafmaß, wenn die üble Nachrede einer breiten Öffentlichkeit zugänglich gemacht wird, also in einem Druckwerk, im Rundfunk oder aber auch im Internet verbreitet wird. Ein Facebook-User wurde beispielsweise (noch nicht rechtskräftig) wegen übler Nachrede verurteilt, weil er einer Politikerin unterstellt hat, ihre Position durch sexuelle Gefälligkeiten erlangt zu haben (DiePresse.com, 2019).

Beleidigungen oder Verspottungen, die ebenfalls nicht von der Meinungsäußerungsfreiheit gedeckt sind und demnach unter Strafe stehen, sind besonders oft in Diskussionsforen oder auf sozialen Medien zu finden. Strafbar ist laut § 115 StGB, wer öffentlich oder vor mehreren Leuten einen anderen beschimpft, verspottet, am Körper misshandelt oder mit einer körperlichen Misshandlung bedroht



(Rechtsinformationssystem Bundeskanzleramt Österreich, 2018b). Die Grenze zwischen einem kritischen, aber zulässigen Werturteil und einer strafbaren Beleidigung ist sehr schwer zu ziehen. Der Oberste Gerichtshof (OGH) hat in einem Fall geurteilt, dass ein Politiker, bei dem die Grenze der zulässigen Kritik zwar weiter gezogen wird als bei einer Privatperson, dennoch nicht grundlos als „Arsch“ beschimpft werden darf (OGH, 2017). In einem anderen Fall hat der EGMR geurteilt, dass ein Politiker von einem Journalisten durchaus als „Trottel“ bezeichnet werden darf, wenn es sich um eine verhältnismäßige Reaktion auf eine provokative Aussage des Politikers handelt (EGMR, 1997).

Eine Verhetzung liegt laut § 283 StGB vor, wenn eine diskriminierende Äußerung gegen eine geschützte Personengruppe getätigt wird. Strafbar ist demnach, „wer öffentlich auf eine Weise, dass es vielen Menschen zugänglich wird, zu Gewalt gegen eine Kirche oder Religionsgemeinschaft oder eine andere nach den vorhandenen oder fehlenden Kriterien der Rasse, der Hautfarbe, der Sprache, der Religion oder Weltanschauung, der Staatsangehörigkeit, der Abstammung oder nationalen oder ethnischen Herkunft, des Geschlechts, einer körperlichen oder geistigen Behinderung, des Alters oder der sexuellen Ausrichtung definierte Gruppe von Personen oder gegen ein Mitglied einer solchen Gruppe ausdrücklich wegen der Zugehörigkeit zu dieser Gruppe auffordert oder zu Hass gegen sie aufstachelt“ (Rechtsinformationssystem Bundeskanzleramt Österreich, 2018c). Zudem werden laut § 283 Abs. 1 Z. 2 StGB auch Äußerungen bzw. Beschimpfungen unter Strafe gestellt, die darauf abzielen, die Menschenwürde dieser geschützten Personengruppe zu verletzen. Wer Material, das zu Hass und Gewalt auffordert oder Hass und Gewalt indirekt fördert, einer breiten Öffentlichkeit zugänglich macht, wird nach § 283 Abs. 4 mit einer Freiheitsstrafe bis zu einem Jahr oder mit einer Geldstrafe bis zu 720 Tagessätzen bestraft. Viele hasserfüllte Äußerungen können allerdings nicht nach § 283 StGB bestraft werden, da oft nicht nachgewiesen werden kann, dass die Tat vorsätzlich begangen wurde (Brodnig, 2016). Vorsätzlich bedeutet in diesem Fall, dass die Täterin oder der Täter es zumindest für möglich gehalten und sich damit abgefunden haben muss, dass durch die getätigte Äußerung eine geschützte Person oder Personengruppe in ihrer Menschenwürde verletzt wird bzw. zu Schaden kommt (Rechtsinformationssystem Bundeskanzleramt Österreich, 2015b).

Ein weiteres Gesetz, das im Spannungsverhältnis zur Meinungsfreiheit steht, ist das Verbotsgesetz (VG). Laut § 3h VG wird eine Person bestraft, wenn sie öffentlich „den nationalsozialistischen Völkermord oder andere nationalsozialistische Verbrechen gegen die Menschlichkeit leugnet, gröblich verharmlost, gutheißt oder zu rechtfertigen sucht.“

(Rechtsinformationssystem Bundeskanzleramt Österreich, 2019d). Personen, die im Internet derartige Äußerungen tätigen, werden somit nicht von der Meinungsäußerungsfreiheit geschützt.

Neben den bisher genannten Bestimmungen aus dem Strafgesetzbuch kommen in Österreich auch Sonderbestimmungen des Mediengesetzes (MedienG) zur Anwendung. Wird beispielsweise der objektiver Tatbestand der üblen Nachrede, Beschimpfung, Verspottung oder Verleumdung in einem (elektronischen) Medium verwirklicht, so steht der betroffenen Person nach § 6 MedienG eine Entschädigung für die erlittene Kränkung zu (Rechtsinformationssystem Bundeskanzleramt Österreich, 2018d)

Wie bereits erwähnt, werden im Internet viele rechtswidrige Inhalte anonym verbreitet. Um die Rechtsverfolgung in diesen speziellen Fällen zu erleichtern, sind Host-Provider in Österreich zur Herausgabe von Userdaten (Name, Adresse) und zur Löschung dieser Inhalte auf ihrer Plattform verpflichtet (Windhager, 2016). Laut § 16 ECG ist ein Host-Provider verpflichtet, rechtswidrige Inhalte von seiner Plattform zu entfernen, sobald er Kenntnis von diesen Inhalten erlangt. Tut er dies nicht, ist er für diese rechtswidrigen Inhalte haftbar (Rechtsinformationssystem Bundeskanzleramt Österreich, 2015c).

### 2.5.3. Gesetzliche Lage in Deutschland

In Deutschland ist die Meinungsäußerungsfreiheit sowohl durch das Grundgesetz (GG) als auch die Europäische Menschenrechtskonvention gewährleistet. Laut Art. 5 Abs. 1 GG hat jeder „das Recht, seine Meinung in Wort, Schrift und Bild frei zu äußern und zu verbreiten und sich aus allgemein zugänglichen Quellen ungehindert zu unterrichten. Die Pressefreiheit und die Freiheit der Berichterstattung durch Rundfunk und Film werden gewährleistet. Eine Zensur findet nicht statt.“ (Deutscher Bundestag, o.J.). Wie in Österreich, sind auch in Deutschland Äußerungen, die nicht von der Meinungsäußerungsfreiheit geschützt sind, im Strafgesetzbuch (StGB) zu finden. Zu den strafbaren Delikten zählen unter anderem die Volksverhetzung, Beleidigung oder üble Nachrede.

Laut § 111 StGB ist es verboten, öffentlich zu Straftaten aufzurufen (dejure.org, o.J.a). Ein Facebook-User wurde demnach beispielsweise verurteilt, weil er die Bundeskanzlerin „öffentlich steinigen“ lassen wollte (Faz.net, 2016).

Das Äquivalent vom österreichischen Tatbestand der „Verhetzung“ ist in Deutschland die „Volksverhetzung“. Strafbar ist laut § 130 Abs. 1 StGB, wer „in einer Weise, die geeignet ist, den öffentlichen Frieden zu stören gegen eine nationale, rassistische, religiöse oder durch ihre ethnische Herkunft bestimmte Gruppe, gegen Teile der Bevölkerung oder gegen einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung zu Hass aufstachelt, zu Gewalt- oder Willkürmaßnahmen auffordert oder die Menschenwürde anderer dadurch angreift, dass er eine vorbezeichnete Gruppe, Teile der Bevölkerung oder einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung beschimpft, böswillig verächtlich macht oder verleumdet“ (dejure.org, o.J.b). Aus § 130 Abs. 1 StGB geht hervor, dass in Deutschland (im Gegensatz zu Österreich), Personen aufgrund ihrer Weltanschauung, ihres Geschlechts, einer körperlichen oder geistigen Behinderung, des Alters oder der sexuellen Ausrichtung nicht geschützt werden. In Deutschland wird ein Verstoß gegen § 130 Abs. 1 StGB mit einer Freiheitsstrafe von bis zu fünf Jahren schärfer bestraft als in Österreich, wo der Täterin oder dem Täter nur eine Freiheitsstrafe von bis zu zwei Jahren droht. Wird hingegen eine Schrift verbreitet, die zu Hass aufstachelt, zu Gewalt auffordert oder die Menschenwürde von Personen einer geschützten Gruppe angreift, kann laut § 130 Abs. 2 StGB entweder mit einer Freiheitsstrafe bis zu drei Jahren oder mit einer Geldstrafe bestraft werden (dejure.org, o.J.b). Für einen Facebook-Post, in dem der Pegida-Gründer Geflüchtete als „Gelumpe“, „Viehzeug“ und „Dreckspack“ beschimpft und dadurch ihre Menschenwürde verletzt hat, wurde vom Gericht eine Geldstrafe von 9.600 Euro verhängt (Zeit Online, 2016).

In Deutschland ist es wie auch in Österreich verboten, Handlungen, die während des Nationalsozialismus begangen wurden, zu billigen, zu leugnen oder zu verharmlosen. Diese Bestimmung ist allerdings nicht in einem eigenen Gesetz, sondern in § 130 Abs. 3 verankert und wird mit einer Freiheitsstrafe von bis zu fünf Jahren oder mit einer Geldstrafe bestraft (dejure.org, o.J.b).

Weiters sind laut § 131 des deutschen Strafgesetzbuchs auch öffentliche Gewaltdarstellungen verboten. Wer unmenschliche Gewalttätigkeiten verbreitet und sie verherrlicht oder verharmlost, wird mit einer Freiheitsstrafe von bis zu einem Jahr oder mit einer Geldstrafe bestraft (dejure.org, o.J.c).

Die wohl am häufigsten begangene Straftat im Internet ist die Beleidigung, die nach § 185 StGB mit einer Freiheitsstrafe von bis zu einem Jahr oder mit einer Geldstrafe

bestraft werden kann (dejure.org, o.J.d). Ein Facebook-User wurde beispielsweise zu einer Geldstrafe von 1.920 Euro verurteilt, weil er eine grüne Politikerin als „linksfaschistische Sau“ bezeichnet hat (Faz.net, 2017).

Von der freien Meinungsäußerung ausgenommen ist in Deutschland auch die üble Nachrede. Laut § 186 StGB macht sich strafbar, „wer in Beziehung auf einen anderen eine Tatsache behauptet oder verbreitet, welche denselben verächtlich zu machen oder in der öffentlichen Meinung herabzuwürdigen geeignet ist“ (dejure.org, o.J.e). Strafbar sind allerdings nur falsche Behauptungen. Wahre Behauptungen sind weiterhin von der Meinungsäußerungsfreiheit geschützt.

Wer Lügen mit voller Absicht verbreitet, macht sich wegen Verleumdung strafbar. Laut § 187 StGB ist strafbar, wer „wider besseres Wissen in Beziehung auf einen anderen eine unwahre Tatsache behauptet oder verbreitet, welche denselben verächtlich zu machen oder in der öffentlichen Meinung herabzuwürdigen oder dessen Kredit zu gefährden geeignet ist“ (dejure.org, o.J.f).

Strafbar ist in Deutschland außerdem nach § 240 StGB, wer eine Person mit Gewalt oder durch eine Drohung zu einer Handlung, Duldung oder Unterlassung nötigt (dejure.org, o.J.g) und nach § 241 StGB, wer eine Person bedroht (dejure.org, o.J.h).

Am 1. Oktober 2017 ist ein Gesetz in Kraft getreten, das die Betreiber gewinnorientierter sozialer Netzwerke mit über zwei Millionen Nutzerinnen und Nutzern unter anderem verpflichtet, offensichtlich rechtswidrige Inhalte wie etwa Volksverletzung, Beleidigung oder Verleumdung innerhalb von 24 Stunden nach Eingang einer Beschwerde zu entfernen (BGBl. I, 2007). Bei Verstößen gegen das Gesetz drohen Unternehmen wie Facebook, YouTube oder Twitter Geldstrafen in Höhe von bis zu 50 Millionen Euro (Deutscher Bundestag, 2017). Das „Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken“, auch „Netzwerkdurchsetzungsgesetz“ (NetzDG) genannt, wurde ziemlich kontrovers aufgenommen. Zwar ist es durchaus ein wichtiger Schritt zur Bekämpfung von hasserfüllten Inhalten im Netz, dennoch haben Kritiker Bedenken, dass sich das Gesetz nicht mit der Meinungsäußerungsfreiheit vereinbaren lässt (Wissenschaftliche Dienste des Deutschen Bundestages, 2017). Ein weiterer Kritikpunkt besteht darin, dass die Entscheidung, ob ein Posting rechtswidrig ist oder nicht, von Unternehmen und nicht von unabhängigen Richtern getroffen wird (Schnellenbach, 2018).

## 2.5.4. Gesetzliche Lage in den Vereinigten Staaten

In den Vereinigten Staaten, ist die Meinungs- bzw. Redefreiheit im 1. Zusatzartikel (First Amendment) zur Verfassung geregelt. Der 1. Verfassungszusatz besagt unter anderem, dass der Kongress keine Gesetze erlassen darf, die die Rede- und Pressefreiheit einschränken (National Constitution Center, o.J.). Der Meinungsfreiheit wird in den Vereinigten Staaten ein großer Stellenwert beigemessen, da sie einerseits den Austausch unterschiedlicher Meinungen erleichtert (Tsesis, 2009) und andererseits Personen ermöglicht, sich an Diskussionen zu beteiligen, ohne Konsequenzen befürchten zu müssen (Amerika Dienst, 2016).

Obwohl die Vereinigten Staaten große Verfechter der Meinungs- bzw. Redefreiheit sind, hat der Supreme Court, der Oberste Gerichtshof der Vereinigten Staaten, einige Kategorien von Äußerungen („unprotected speech“) bestimmt, die nicht vom 1. Verfassungszusatz geschützt werden (Bezemek, 2015). Demnach sind unter anderem Obszönitäten, „Kampfwörter“ („fighting words“), ernste Drohungen und Verleumdungen verfassungswidrig und nicht vom First Amendment geschützt (Ruane, 2014). Im Fall *Chaplinsky v. New Hampshire* hat Richter Frank Murphy die Grenze zwischen Meinungsfreiheit und HOC wie folgt gezogen: „There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which has never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or 'fighting' words—those which by their very utterance inflict injury or tend to incite an immediate breach of the peace.“ (LexisNexis, 1942).

Im Gegensatz zu Deutschland, wo Unternehmen wie Facebook, Twitter & Co für rechtswidrige Inhalte haften, wenn sie nicht rechtzeitig gelöscht werden, werden in den Vereinigten Staaten Internet-Provider durch § 230 des Communication Decency Act (CDA) geschützt, indem sie nicht für die Inhalte auf ihren Plattformen haftbar gemacht werden. § 230 CDA besagt einerseits (1) „No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.“ sowie (2) „No provider or user of an interactive computer service shall be held liable on account of any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or any action

taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph” (Legal Information Institute, o.J.). In den Vereinigten Staaten sind somit Unternehmen wie Facebook, Twitter oder YouTube nicht für rechtswidrige Inhalte auf ihren Plattformen verantwortlich.

### 2.5.5. Gesetzliche Lage in Großbritannien

In Großbritannien ist die freie Meinungsäußerung in Art. 10 Abs. 1 des Human Rights Act 1998 wie folgt geregelt: „Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.” (legislation.gov.uk, 2004). Laut Abs. 2 kann die freie Meinungsäußerung allerdings aus wichtigen Gründen (z.B. nationale Sicherheit oder zum Schutz des Rufs oder der Rechte anderer) eingeschränkt werden. Weil Großbritannien ein Mitglied des Europarats ist, ist die freie Meinungsäußerung auch durch die Europäische Menschenrechtskonvention gewährleistet.

In England, Wales und Schottland wird HOC zunächst durch den Public Order Act 1986 (POA) reguliert. Aus dem dritten Teil des POA geht hervor, dass rassistischer Hass („racial hatred“) gegen eine Personengruppe aufgrund ihrer Hautfarbe, Abstammung („race“), Nationalität, ethnischen oder nationalen Herkunft verboten ist und mit einer Freiheitsstrafe von bis zu sieben Jahren oder einer Geldstrafe bestraft werden kann (legislation.gov.uk, 2001). 2006 wurde durch § 1 des Racial and Religious Hatred Act auch das Schüren von Hass aus religiösen Gründen unter Strafe gestellt (legislation.gov.uk, 2007). Seit dem Criminal Justice and Immigration Act 2008 ist außerdem die Aufstachelung zu Hass aufgrund der sexuellen Orientierung strafbar (legislation.gov.uk, 2010).

In § 18 POA steht: „A person who uses threatening, abusive or insulting words or behaviour, or displays any written material which is threatening, abusive or insulting, is guilty of an offence if he intends thereby to stir up racial hatred, or having regard to all the circumstances racial hatred is likely to be stirred up thereby.” (legislation.gov.uk, 2017a).

Der Public Order Act 1986 verbietet außerdem nach § 4 drohende, missbräuchliche oder beleidigende Worte oder Verhaltensweisen, die bei einer Person Angst auslösen oder

Gewalt provozieren und nach § 4A Worte oder Verhaltensweisen, die eine Person vorsätzlich belästigen, verängstigen oder beunruhigen (legislation.gov.uk, 2017b).

Im Malicious Communications Act 1988 (MCA) sind weitere Vorschriften bezüglich HOC geregelt. Laut § 1 MCA sind grob beleidigende Äußerungen verboten, die geeignet sind, Ärger und Angst zu verursachen (legislation.gov.uk, 2015a). § 127 des Communications Act 2003 verbietet weiters jede grob beleidigende, obszöne oder drohende Nachricht, die mittels eines öffentlichen elektronischen Kommunikationsnetzes gesendet wird (legislation.gov.uk, 2015b).

Großbritannien hat im Vergleich zu Deutschland kein Gesetz, welches Unternehmen verpflichtet, rechtswidrige Inhalte innerhalb einer bestimmten Frist zu löschen. Dennoch scheint es Befürworter für ein solches Gesetz zu geben. Der konservative Abgeordnete Andrew Percy ist beispielsweise der Meinung, dass die sozialen Medien zu einem gesetzlosen Raum geworden sind und spricht sich deshalb klar für ein System mit Fristen und Geldstrafen wie in Deutschland aus (The Telegraph, 2018). Einige Abgeordnete gehen sogar noch einen Schritt weiter und fordern, dass Führungskräfte von Unternehmen wie Facebook oder Twitter für schädliche Inhalte auf ihren Plattformen zur Verantwortung gezogen werden müssen (Financial Times, 2019).





### III. Die Verantwortung von Unternehmen und ihre Maßnahmen gegen HOC

Obwohl Harmful Online Communication für die Betroffenen erwiesenermaßen äußerst schädlich ist, wird HOC in vielen Ländern zum Teil noch durch die freie Meinungsäußerung geschützt. Wenn die nationalen und internationalen rechtlichen Institutionen nicht umfassend gegen HOC vorgehen können, stellt sich die Frage, ob es eine andere Möglichkeit gibt, HOC einzudämmen.

Sowohl in der Gesellschaft als auch in der Politik oder Wissenschaft wird viel darüber diskutiert, wer für HOC verantwortlich ist und wer am effizientesten dagegen vorgehen kann. Alexander Brown (2018), der am Institut für Politik, Philosophie, Sprache und Kommunikationswissenschaft an der University of East Anglia lehrt, ist der Meinung, dass die Politik einen eingeschränkten Handlungsspielraum hat. Daher sollten Internetunternehmen, die ihre Plattformen der Öffentlichkeit zur Verfügung stellen, für die Regulierung von HOC zuständig sein. Facebook, Twitter & Co erleichtern und fördern zwar den digitalen Dialog und die Verbreitung von Informationen, dennoch werden ihre Plattformen auch genutzt, um Hassbotschaften und andere schädliche und illegale Inhalte zu verbreiten. Wissenschaftler wie Brown oder Cohen-Almagor (2017) sind sich deswegen einig, dass diese Unternehmen eine soziale Verantwortung gegenüber der Öffentlichkeit, der sie dienen, haben.

Den großen Internetkonzernen sollte bewusst sein, dass sie eine entscheidende Rolle bei der Verhinderung bzw. Reduzierung von HOC spielen, da sie in der Lage sind, entsprechende Maßnahmen gegen den Hass auf ihren Plattformen zu treffen (Banks, 2010). Durch ihre Gatekeeper-Funktion können sie nämlich entscheiden, welche Informationen durchgelassen werden und welche nicht. Dadurch haben sie eine enorme Macht, mit der aber auch eine entsprechende Verantwortung einher geht (Cohen-Almagor, 2017). Kommunikationswissenschaftler Damian Trilling erklärt in einem Interview mit *derStandard.at* (2016b), dass Facebook ein technologisches Unternehmen ist, das journalistische Funktionen übernimmt, weil es durch seine Algorithmen Informationen selektiert und entsprechend aufbereitet. Aus diesem Grund sollte Facebook laut Trilling auch eine journalistische Verantwortung haben. Ingrid Brodnig (2016) ist ebenfalls der Ansicht, dass Facebook eine besondere Verantwortung zukommt und

kritisiert, dass die Verantwortlichen bestreiten, einen Einfluss auf die Informationsselektion zu haben.

Außerdem sollten neben den großen Internetkonzernen auch Onlinemedien, Webseitenbetreiber und Social-Media-Verantwortliche für die Inhalte auf ihren Diskussionsplattformen verantwortlich sein. Sie können nämlich einen maßgeblichen Einfluss auf den dort vorherrschenden Umgangston haben (Brodnig, 2016). Unternehmen, die Online-Plattformen betreiben, können nämlich als Selbstregulierer fungieren und Regeln und Richtlinien erstellen, an die sich die Userinnen und User halten müssen (Brown, 2018). Durch Nutzungsbedingungen, Verhaltenskodizes oder Netiquetten können sie kommunizieren, dass sie sich für ein respektvolles und konstruktives Diskussionsklima einsetzen und somit aktiv gegen aggressive, schädliche oder illegale Inhalte vorgehen. Zu den weiteren Maßnahmen, die die Unternehmen treffen können, um gegen HOC vorzugehen, zählen unter anderem das Moderieren von Diskussionsplattformen, das Verbot von anonymen Postings oder das Löschen von rechtswidrigen Inhalten. Diese und weitere Maßnahmen werden in den folgenden Kapiteln vorgestellt.

## 3.1. Lösungsvorschläge für den Umgang mit HOC

### 3.1.1. Verhaltensrichtlinien und Forenregeln

Wenn ein Land wie die Vereinigten Staaten keine ausreichenden Maßnahmen oder rechtlichen Konsequenzen gegen HOC setzt, sind die Unternehmen dazu angehalten, selbstständig zu handeln. Unternehmen wie Facebook, Twitter, YouTube oder Microsoft haben sich beispielsweise bereits zur Bekämpfung illegaler Hassrede im Internet verpflichtet. In Zusammenarbeit mit der Europäischen Kommission haben die IT-Unternehmen einen Verhaltenskodex entwickelt, der darauf abzielt, die Verbreitung von unerwünschten Hassbotschaften auf Online-Plattformen zu unterbinden. Als eine notwendige Maßnahme zur Regulierung von HOC wurde hierbei das Erstellen von Community-Richtlinien erachtet. Die Verhaltensrichtlinien, die für alle Userinnen und User gelten, sollen unter anderem gewalttätiges und hasserfülltes Verhalten und andere unerwünschte Inhalte verbieten. Die Unternehmen haben sich zudem verpflichtet, ihre Userinnen und User über ihre Verhaltensrichtlinien aufzuklären (Code of Conduct on Countering Illegal Hate Speech Online, 2016).

In den Verhaltensrichtlinien ist typischerweise geregelt, welche Inhalte auf den Online-Plattformen verboten sind, wie sich Userinnen und User verhalten sollen und welche Konsequenzen bei einem Verstoß gegen die Richtlinien drohen. Aus den aktuellen Gemeinschaftsstandards von Facebook geht beispielsweise hervor, dass Inhalte entfernt werden, wenn sie in der realen Welt Schäden verursachen können. Darunter fallen sowohl gewalttätige und kriminelle als auch anstößige Inhalte. In seinen Gemeinschaftsstandards fordert Facebook einen respektvollen Umgang sowie ein verantwortungsbewusstes Verhalten. Zu den Verantwortungen der Userinnen und User zählt beispielsweise das Melden von Inhalten, die gegen die Gemeinschaftsstandards verstoßen. Zuletzt betont Facebook in seiner Richtlinie auch, dass Verstöße unterschiedliche Konsequenzen haben können, die von einer Verwarnung bis hin zur Deaktivierung von Profilen reichen (Facebook, 2019).

Twitter hat ebenfalls Regeln für die Nutzung seiner Plattform erstellt. In den „Twitter Rules“, die regelmäßig ergänzt und angepasst werden, steht unter anderem, dass die sexuelle Ausbeutung von Kindern, Gewalt, Missbrauch, Belästigung und hasserfüllte

Inhalte verboten sind. Das Unternehmen begründet die Notwendigkeit von Verboten damit, dass gewisse Inhalte einen freien und sicheren öffentlichen Diskurs verhindern (Twitter, 2019).

YouTube appelliert in seinen Community-Richtlinien ebenfalls an das Verantwortungsbewusstsein der Userinnen und User. Sie werden gebeten, unter anderem sexuelle, schädliche, gefährliche, hasserfüllte, gewalttätige oder grausame Inhalte zu melden. YouTube warnt außerdem ausdrücklich vor Verstößen gegen die Community-Richtlinien und droht mit der Entziehung von Privilegien und Kontokündigungen (YouTube, 2019).

Neben den großen Konzernen wie Facebook, Twitter oder YouTube verpflichten sich mittlerweile auch Webseitenbetreiber, Unternehmen mit öffentlichen Profilen auf sozialen Netzwerken oder Onlinemedien, die ihren Leserinnen und Lesern Diskussionsplattformen zur Verfügung stellen, zur Regulierung von HOC. Die Onlinezeitung *derStandard.at* hat beispielsweise besonders ausführliche Forenregeln festgelegt, an die sich alle Personen, die Beiträge kommentieren wollen, halten müssen. Darin werden ein respektvoller Umgang, sachliche Argumentationen und die Einhaltung von Gesetzen und Rechtsvorschriften gefordert. Außerdem werden Diskriminierungen und Diffamierungen ausdrücklich verboten. Bei Missachtung der Forenregeln behält sich *derStandard.at* das Recht vor, Beiträge zu entfernen oder Userinnen und User zu sperren (derStandard.at, 2018).

### 3.1.2. Registrierung auf Online-Plattformen

Viele Betreiber von Online-Plattformen bzw. Diskussionsforen überlegen sich, wie sie destruktive, hasserfüllte und illegale Inhalte in ihren Foren verhindern bzw. einschränken können. Neben dem Erstellen und aktiven Kommunizieren von Verhaltensrichtlinien werden auch andere Maßnahmen getroffen, um gegen HOC vorzugehen. Die *Huffington Post* verbietet beispielsweise bereits seit 2013 das Kommentieren von Beiträgen unter einem anonymen Benutzernamen (Landers, 2013).

Wissenschaftler haben untersucht, ob das Verbot von anonymen Kommentaren tatsächlich eine effektive Maßnahme ist, um HOC einzuschränken. Anhand einer Studie wollte Kommunikationswissenschaftler Thomas B. Ksiazek (2015) herausfinden, wie

Betreiber von Nachrichtenwebseiten erfolgreich dazu beitragen können, dass Anfeindungen auf ihren Plattformen eingedämmt werden. Herausgefunden wurde, dass ein höflicheres Diskussionsklima bereits durch einige unkomplizierte Maßnahmen erzielt werden kann. Eine besonders effektive Maßnahme ist die verpflichtete Registrierung auf der Online-Plattform. Dadurch soll verhindert werden, dass Menschen unter dem Deckmantel der Anonymität anstößige und beleidigende Kommentare hinterlassen. Die erleichterte Identifizierbarkeit soll die Hemmschwelle erhöhen und so zu einem insgesamt höflicheren Umgangston verhelfen.

Die österreichische Regierung ist ebenfalls davon überzeugt, dass die Registrierung von Nutzerinnen und Nutzern in Online-Foren eine effektive Maßnahme gegen Hass im Netz ist. Durch das geplante „Bundesgesetz über Sorgfalt und Verantwortung im Netz“, soll der respektvolle Umgang auf Diskussionsplattformen gefördert werden. Außerdem soll durch das geplante Gesetz die Rechtsverfolgung bei rechtswidrigen Kommentaren erleichtert werden. Ähnlich wie beim deutschen „Netzwerkdurchsetzungsgesetz“ sollen außerdem Unternehmen verpflichtet werden, gemeldete Inhalte innerhalb von 24 Stunden zu überprüfen (parlament.gv.at, 2019).

### 3.1.3. Moderation der Online-Plattformen

In einer demokratischen Gesellschaft können Personen uneingeschränkt über kontroverse und polarisierende Themen diskutieren und debattieren. Ihnen werden zum Meinungsaustausch unterschiedliche Plattformen zur Verfügung gestellt, darunter auch Diskussionsforen von Onlinemedien. Dort wo vielseitige Meinungen aufeinandertreffen, ist es umso wichtiger, für ein respektvolles Miteinander und ein faires, sachliches und angenehmes Diskussionsklima zu sorgen. Viele Onlinemedien versuchen bereits anhand von Community- oder Content Guidelines die Qualität der Diskussionen zu fördern, indem sie gewisse Äußerungen oder Inhalte auf ihren Plattformen verbieten. Diese Verbote sind laut Ingrid Brodnig (2016) deshalb besonders wichtig, weil aggressive und beleidigende Kommentare eine sachliche Auseinandersetzung mit dem Diskussionsthema unmöglich machen.

Eine weitere Maßnahme, die für ein verbessertes Diskussionsklima sorgt und deswegen eingesetzt wird, um gezielt gegen HOC vorzugehen, ist die Moderation der Kommentare

auf den Diskussionsplattformen (Ksiazek, 2015). In der Praxis gibt es im Wesentlichen zwei unterschiedliche Vorgehensweisen. Die Unternehmen bzw. Betreiber von Online-Plattformen können Mitarbeiterinnen und Mitarbeiter beauftragen, die Kommentare zu moderieren. Dabei ist die Prä-Moderation von der Post-Moderation zu unterscheiden. Bei der Prä-Moderation werden die Kommentare vor der Veröffentlichung auf ihre Rechtmäßigkeit geprüft. Bei der Post-Moderation beteiligen sich die Moderatorinnen und Moderatoren aktiv an den Diskussionen und schreiten ein, wenn Personen gegen die Richtlinien verstoßen. Darüber hinaus gibt es auch die Möglichkeit, die Userinnen und User in die Moderation miteinzubeziehen (Diakopoulos & Naaman, 2011). Diese Art der Moderation nennt sich „crowdsourced content moderation“ und hat mehrere Funktionen. Einerseits sollen die Userinnen und User die Möglichkeit erhalten, Kommentare positiv oder negativ zu bewerten. Positive Kommentare werden dadurch im Diskussionsforum prominenter platziert, negative Kommentare in den Hintergrund gerückt. Kommentare, die gegen die Richtlinien verstoßen, sind dadurch auch leichter zu lokalisieren. Andererseits kann crowdsourced content moderation auch zu einem zivileren Diskussionsklima führen. Ingrid Brodnig hat in ihrem Buch „Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können“ ein Start-up aus den Vereinigten Staaten vorgestellt, das anhand einer Software „das Gefühl der Unsichtbarkeit im Netz“ bekämpfen sollte (Brodnig, 2016, S. 198). Zwar gibt es das Unternehmen „Civil Comments“ mittlerweile nicht mehr, dennoch ist die Idee dahinter durchaus erwähnenswert. Um ein Kommentar auf einer Diskussionsplattform hinterlassen zu können, mussten Userinnen und User zunächst andere Kommentare in Hinblick auf ihre Qualität und Höflichkeit bewerten. Außerdem mussten sie beurteilen, ob die jeweiligen Kommentare Beleidigungen oder andere persönliche Angriffe beinhalten. Im Anschluss wurden die Userinnen und User gefragt, ob ihr Kommentar zivil bzw. höflich und respektvoll formuliert ist. Erst dann wurden Kommentare zur Veröffentlichung freigegeben (Civil Comments, 2016).

In vielen Medienunternehmen werden unterschiedliche Moderationssysteme eingesetzt. Beim *Standard* kümmern sich beispielsweise zwölf Mitarbeiterinnen und Mitarbeiter um die Foren mit derzeit fast 60.000 aktiven Userinnen und User und um die zehn Millionen verfassten Beiträge pro Jahr (derStandard.at, 2019b). Die Foren werden einerseits prä-moderiert, da alle Beiträge der Userinnen und User zunächst anhand einer Software geprüft und dann erst automatisch veröffentlicht werden. Andererseits werden die Foren durch die aktive Präsenz der Moderatorinnen und Moderatoren auch post-moderiert (derStandard.at, 2016c).

Die Moderation von Kommentaren und anderen Inhalten spielt auch auf den sozialen Netzwerken eine bedeutende Rolle. Facebook setzt hierfür beispielsweise neben künstlicher Intelligenz und maschinellem Lernen auch auf speziell geschulte Mitarbeiterinnen und Mitarbeiter, da die automatische Spracherkennung in manchen Fällen nicht zwischen zulässigen und unzulässigen Äußerungen unterscheiden kann (Davidson, Warmsley, Macy, & Weber, 2017). Weltweit sind um die 15.000 Personen im Einsatz, die die gemeldeten Inhalte auf ihre Rechtmäßigkeit überprüfen (Facebook Newsroom, 2018).

Facebook kann allerdings nur jene Inhalte überprüfen, die von anderen Userinnen und Usern gemeldet wurden. Deswegen ist es wichtig, dass auch Betreiber von Facebook-Seiten und Social-Media-Verantwortliche von Unternehmen, die ein öffentliches Profil betreiben, die Kommentare auf ihren Seiten überwachen bzw. moderieren. Das ist für ein respektvolles Diskussionsklima unerlässlich, da aggressive Kommentare häufig Reaktionen von anderen Userinnen und Usern auslösen. Je mehr Personen auf ein Kommentar reagieren, desto mehr Personen wird es auch angezeigt, da es vom Facebook-Algorithmus als relevant erachtet wird (Brodnig, 2016). So kann es schnell passieren, dass nur noch die destruktiven Kommentare gelesen werden und die konstruktiven Kommentare in der Masse untergehen.

### 3.1.4. Umgang mit HOC-Inhalten

Viele Unternehmen verpflichten sich freiwillig zur Löschung von destruktiven, schädlichen oder sonstigen rechtswidrigen Inhalten. Die Unternehmen kommunizieren durch ihre Nutzungsbedingungen, Community- oder Content Guidelines, welche Inhalte auf ihren Plattformen nicht geduldet werden. In den Richtlinien wird auch geregelt, welche Maßnahmen bei einem Verstoß getroffen werden. Die Maßnahmen reichen dabei je nach Schwere des Verstoßes vom Kürzen oder Löschen des unerwünschten Inhaltes bis hin zur Sperre des Benutzerkontos. In der Netiquette von *Zeit Online* wird beispielsweise genau erklärt, wann und wie die Moderatorinnen und Moderatoren einschreiten. Bei Verstößen gegen die Verhaltensregeln werden die Kommentare entweder gekürzt oder entfernt, je nachdem ob der Beitrag nur teilweise oder zur Gänze regelwidrig ist. Wird ein Kommentar gekürzt oder gelöscht, erscheint an der Stelle eine Anmerkung der Moderatorinnen und Moderatoren. Bei schweren oder wiederholten Verstößen verwehrt

*Zeit Online* den Userinnen und Usern auch den Zugang zur Diskussionsplattform (*Zeit Online*, o.J.).

In einigen Ländern ist die Löschung von rechtswidrigen Beiträgen auf Online-Plattformen nicht optional. Bestimmte Unternehmen bzw. Betreiber von Online-Plattformen sind gesetzlich dazu verpflichtet, rechtswidrige Inhalte innerhalb einer bestimmten Frist zu entfernen. In Deutschland besagt das NetzDG, dass volksverhetzende oder beleidigende Inhalte, sofern sie offensichtlich rechtswidrig sind, innerhalb von 24 Stunden nach Eingang einer Beschwerde entfernt werden müssen (BGBl. I, 2007). In Österreich sind Host-Provider ebenfalls nach § 16 Abs. 1 Z. 2 ECG gesetzlich dazu verpflichtet, offensichtlich rechtswidrige Inhalte nach Kenntnisnahme unverzüglich zu entfernen bzw. den Zugang zu diesen Inhalten zu sperren (Rechtsinformationssystem Bundeskanzleramt Österreich, 2015c).



## IV. Wie Betroffene gegen HOC vorgehen können

Unternehmen und Betreiber von Online-Plattformen sollten ihren Userinnen und Usern unterschiedliche Möglichkeiten anbieten, um auch selbstständig gegen HOC vorgehen zu können. Dazu zählen vor allem technische Möglichkeiten wie das Blockieren von Userinnen und Usern oder das Melden von aggressiven oder anstößigen Inhalten. Darüber hinaus ist es aber auch wichtig, die Userinnen und User darüber zu informieren, wie sie mit HOC umgehen können.

### 4.1. Blockieren und Melden

Im Internet ist häufig die Faustregel „Don't feed the troll“ zu lesen. Laut Ingrid Brodnig (2016) ist eine bewehrte Methode gegen Internettrolle, die dafür bekannt sind, sich in Diskussionsforen oder auf sonstigen Online-Plattformen destruktiv und störend zu verhalten, sie einfach zu ignorieren. Jegliche Reaktion, sei es eine einfache Gegenrede oder das Löschen des Kommentares, sei in dem Fall kontraproduktiv, da Trolle bewusst provozieren wollen.

Es ist allerdings durchaus verständlich, dass es auch Personen gibt, die beleidigende Kommentare oder sonstige anstößige Inhalte nicht ignorieren und tolerieren wollen. Hier rät Ingrid Brodnig (2016) zum Blockieren des Profils der Täterin oder des Täters. Einerseits, um sich selbst zu schützen und andererseits, um in Zukunft nicht mehr mit derartigen Inhalten konfrontiert zu werden.

Offensichtlich rechtswidrige oder gegen die Verhaltensrichtlinien verstoßende Beiträge sollten außerdem zunächst dokumentiert und anschließend gemeldet werden (Brodnig, 2016). Wie bereits erwähnt, sind nämlich die Betreiber von Online-Plattformen in Deutschland oder Österreich nach Kenntnisnahme zur Löschung von rechtswidrigen Inhalten verpflichtet (BGBl. I, 2007; Rechtsinformationssystem Bundeskanzleramt Österreich, 2015c).

## 4.2. Counter Speech

Oft wissen Betroffene oder unbeteiligte Dritte nicht wie sie auf HOC reagieren sollen. Viele Online-Plattformen reagieren bereits auf HOC und blenden entsprechende Inhalte aus oder entfernen sie zur Gänze. Darüber hinaus sollten Unternehmen und Webseitenbetreiber aber auch ihre Userinnen und User ermutigen, aktiv dagegen vorzugehen und HOC nicht einfach hinzunehmen. Eine Möglichkeit des aktiven Handelns ist dabei das Melden von unangemessenen und gegen die Richtlinien verstoßenden Inhalten. Eine weitere Möglichkeit ist die Gegenrede, auch „Counter Speech“ genannt. Hierbei reagieren Betroffene oder unbeteiligte Dritte auf aggressive, beleidigende oder hasserfüllte Äußerungen. Diese aktive Konfrontation erfordert Mut und Überwindung, da Personen in Kauf nehmen müssen, weiteren Beleidigungen ausgesetzt zu werden bzw. ebenfalls zur Zielscheibe gemacht zu werden.

Damit Counter Speech auch seinen Zweck erfüllt, müssen einige Regeln beachtet werden. Die Gegenrede sollte wohl überlegt bzw. sachlich formuliert werden und stichhaltige Argumente beinhalten. Die Täterinnen und Täter sollen allerdings nicht belehrt oder zurechtgewiesen werden. Das primäre Ziel ist, dass die destruktiven Kommentare an Gewicht verlieren (Saferinternet.at, o.J.). Effektive Gegenreden müssen allerdings nicht zwangsläufig gute Gegenargumente beinhalten. Sie können auch durchaus humorvoll formuliert sein, da Humor zur Deeskalation beitragen kann (Brodnig, 2016). Ein oft zitiertes Beispiel für die effektive Wirkung von Humor ist die Reaktion von James Blunt auf eine Beleidigung auf Twitter. Ein User schrieb in einem Tweet, dass der Sänger aussehe wie sein linker Hoden. James Blunt nahm es mit Humor und antwortete, dass er besser einen Arzt aufsuchen sollte (Blunt, 2014).

Counter Speech ist eine wichtige Strategie gegen HOC, weil sie einerseits versucht Grenzen zu setzen, indem beleidigende oder anstößige Äußerungen kommentiert und nicht einfach ignoriert und hingenommen werden. Andererseits kann die Gegenrede einer Person auch andere Personen dazu ermutigen, sich gegen derartige Äußerungen auszusprechen. Klaus Schwertner, Generalsekretär der Caritas, hat eindrucksvoll bewiesen, wie effektiv Counter Speech sein kann. Als 2018 die kopftuchtragende Mutter und der Vater des „Wiener Neujahrsbabys“ in den sozialen Medien mit unzähligen fremdenfeindlichen Kommentaren konfrontiert wurden, hat er die Gegeninitiative „#flowerrain“ ins Leben gerufen. Schwertner hat in einem Facebook-Post, der mittlerweile 16.591 Mal geteilt und 32.547 Mal kommentiert wurde, erwähnt, dass er ein Buch mit

„Glückwünschen, netten Worten und Willkommensnachrichten“ an die Familie übergeben möchte (Schwertner, 2018). Innerhalb kürzester hat diese Aktion für eine Welle von positiven Kommentaren von anderen Userinnen und Usern gesorgt.

### 4.3. Rechtliche Schritte

Wenn Personen im Internet beleidigt oder verleumdet werden und die Plattform-Betreiber ihrer Verpflichtung zur Löschung nicht nachkommen, sollten sie selbstständig rechtliche Schritte einleiten. In Österreich können Betroffene unter anderem klagen, wenn sie in ihren Persönlichkeitsrechten verletzt werden (Brodnig, 2016). § 6 MedienG gewährt den Betroffenen bei übler Nachrede, Beschimpfung, Verspottung und Verleumdung eine Entschädigung für die erlittene Kränkung (Rechtsinformationssystem Bundeskanzleramt Österreich, 2018d).

Darüber hinaus können die Betroffenen bei Ehrverletzungen auch zivilrechtlich vorgehen. Neben dem MedienG gewährt auch § 1330 des Allgemeinen Bürgerlichen Gesetzbuches (ABGB) materiellen Schadenersatz, wenn „jemandem durch Ehrenbeleidigung ein wirklicher Schade oder Entgang des Gewinnes verursacht worden ist“ (Rechtsinformationssystem Bundeskanzleramt Österreich, 2015d). Diese Bestimmung ist besonders wichtig, weil viele Beleidigungen im Netz bestimmte Kriterien nicht erfüllen, weshalb die Täterinnen und Täter nicht nach dem StGB bestraft werden können (ZARA - Zivilcourage & Anti-Rassismus-Arbeit, 2018).

# V. Empirie

## 5.1. Methode

In diesem Kapitel wird zunächst die für diese Forschungsarbeit verwendete empirische Methode, nämlich die der Inhaltsanalyse, vorgestellt. Außerdem werden unter anderem das methodische Vorgehen, die Stichprobe, das Erhebungsinstrument und das Codebuch näher beschrieben.

Die Inhaltsanalyse ist die Methode, die in den Forschungen der Kommunikationswissenschaft am häufigsten angewendet wird (Brosius, Koschel, & Haas, 2008). Laut Werner Früh ist die Inhaltsanalyse „eine empirische Methode zur systematischen und intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen, meist mit dem Ziel einer darauf gestützten interpretativen Inferenz auf mitteilungsexterne Sachverhalte“ (Früh, 2017, S. 29).

Bei einer Inhaltsanalyse werden zuvor festgelegte formale und inhaltliche Merkmale großer Textmengen erfasst und analysiert (Brosius, Koschel, & Haas, 2008). In der vorliegenden Arbeit werden unterschiedliche Unternehmensrichtlinien aus vier Ländern anhand eines nachvollziehbaren Codebuchs systematisch analysiert. Hierfür müssen laut Brosius, Koschel & Haas (2008) folgende Schritte durchgeführt werden: zunächst müssen der Fragestellung entsprechende, relevante Kategorien gebildet werden. Nach der Kategorienbildung folgt die Erprobung und Überprüfung des Codebuchentwurfs. 10% des zu untersuchenden Materials werden in einem späteren Schritt im Rahmen des Pretests probecodiert. In dieser Phase wird das Codebuch kontrolliert und gegebenenfalls bei etwaigen Unklarheiten auch adaptiert. Mit der eigentlichen Codierung kann erst begonnen werden, wenn intersubjektive Nachvollziehbarkeit gegeben ist. Intersubjektiv nachvollziehbar ist eine Inhaltsanalyse dann, wenn das Messinstrument derart zuverlässig ist, dass die Inhaltsanalyse von einem anderen Codierer fehlerfrei wiederholt werden kann. Die Zuverlässigkeit des Codebuchs wird anhand eines Reliabilitätstest gemessen. Um die Zuverlässigkeit der Ergebnisse zu garantieren, sollte auch während der laufenden Codierung ein weiterer Reliabilitätstest durchgeführt werden. Der letzte Schritt der Inhaltsanalyse besteht aus der Auswertung und Analyse des gesamten Untersuchungsmaterials.

### 5.1.1. Methodische Vorgehensweise

Bei dieser Arbeit handelt es sich um die Erweiterung einer bestehenden englischsprachigen Forschungsarbeit von Univ.-Prof. Dr. Sabine Einwiller (Universität Wien) und Assoc. Prof. Dr. Sora Kim (The Chinese University of Hong Kong). Ziel der Forschungsarbeit von Einwiller und Kim war es unter anderem herauszufinden, wie diverse Unternehmen (u.a. Blog-Hosting Site Betreiber oder Medienunternehmen) mit HOC umgehen und welche Richtlinien sie kommunizieren, um HOC entgegenzuwirken. Um dieser Frage nachzugehen, wurden diverse Richtlinien (Nutzungsbedingungen, Community Guidelines, Content Guidelines und Reporting Guidelines) von Unternehmen aus Österreich, Deutschland, Großbritannien, den Vereinigten Staaten, China, Japan und Südkorea inhaltsanalytisch untersucht. Außerdem wurden Interviews mit Vertreterinnen und Vertretern von Organisationen, die für das Community und/oder Social Media Management verantwortlich sind, geführt. Anhand dieser Interviews sollte herausgefunden werden, wie Unternehmen Richtlinien in Bezug auf HOC in der Praxis anwenden und welche Erfahrungen sie damit machen (Einwiller & Kim, 2018).

Für diese Forschungsarbeit wurde das Untersuchungsmaterial von Einwiller und Kim erneut herangezogen. Allerdings wurden hier nur jene Richtlinien von Unternehmen aus Österreich (AU), Deutschland (GE), den Vereinigten Staaten (US) und Großbritannien (GB) systematisch anhand des unveröffentlichten Codebuchs von Einwiller und Kim analysiert. Da sich diese Arbeit wesentlich detaillierter mit den unterschiedlichen Richtlinien der Unternehmen befasst, wurde das bestehende Codebuch um weitere Variablen ergänzt.

### 5.1.2. Stichprobe

Auf der Grundlage von Web-Traffic-Daten wurden die drei führenden Unternehmen der vier Ländern (Österreich, Deutschland, den Vereinigten Staaten und Großbritannien) aus folgenden Kategorien ausgewählt: Web-Portal-Webseiten (Web Portal Sites), Blog-Hosting-Webseiten (Blog Hosting Sites), Community-Webseiten (Community Sites), E-Commerce-Webseiten (E-Commerce Sites) und Empfehlungsportale (Recommendation Portals). Außerdem wurden pro Land acht Webseiten großer Nachrichtenmedien (Major News Media Sites), fünf der beliebtesten Social Media Webseiten (Most Popular Social Network Sites) und weitere zehn große Unternehmen (Large non-Internet companies: Website and/or SNS) ausgewählt.

Somit wurden pro Land Richtlinien von 38 Unternehmen analysiert. Die ausgewählten Unternehmen aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien werden in den Tabellen 1 bis 4 dargestellt.

AU		
3	Web Portal Sites	Yahoo, GMX, Web.de
3	Blog Hosting Sites	WordPress, Tumblr, Jimdo
3	Community Sites	Gutefrage.net, Ichkoche.at, Chefkoch.de
3	E-Commerce Sites	Amazon, eBay, ASOS
3	Recommendation Portals	DocFinder, TripAdvisor, HolidayCheck
8	Major News Media Sites	Oe24, DerStandard, Kronen Zeitung, Vorarlberger Nachrichten, Kurier, Salzburger Nachrichten, Kleine Zeitung, Die Presse
5	Most Popular Social Network Sites (SNS)	YouTube, Facebook, Instagram, Reddit, Twitter
10	Large non-Internet companies: Website	OMV Group, Voest Alpine, Red Bull
	Large non-Internet companies: SNS	Billa, BMW, Novomatic, Benteler
	Large non-Internet companies: Website and SNS	Hofer, Spar, ÖBB
38		

Tabelle 1: Ausgewählte Unternehmen aus Österreich

GE		
3	Web Portal Sites	Web.de, Yahoo, GMX
3	Blog Hosting Sites	Blogger, WordPress, Tumblr
3	Community Sites	Gutefrage.net, Imgur, Chefkoch.de
3	E-Commerce Sites	Amazon, eBay, ASOS
3	Recommendation Portals	Yelp, TripAdvisor, Ciao.de
8	Major News Media Sites	Spiegel, Bild, CHIP, Focus, Die Welt, Die Zeit, Frankfurter Allgemeine Zeitung, Süddeutsche Zeitung
5	Most Popular Social Network Sites (SNS)	Facebook, Twitter, LinkedIn, XING, YouTube
10	Large non-Internet companies: SNS	Allianz, Siemens Home, Munich Re, Bayer Group, Deutsche Telekom, BMW Group, SAP, Fresenius, Mercedes Benz, Volkswagen
38		

Tabelle 2: Ausgewählte Unternehmen aus Deutschland



US		
3	Web Portal Sites	Yahoo, MSN, AOL
3	Blog Hosting Sites	Weebly, WordPress, Blogger
3	Community Sites	Stack Exchange, Allrecipes, wikiHow
3	E-Commerce Sites	Amazon, Walmart, Macy's
3	Recommendation Portals	Yelp, TripAdvisor, Foursquare
8	Major News Media Sites	USA Today, The New York Times, New York Post, Los Angeles Times, NewsDay, Chicago Tribune, Daily News, The Washington Post
5	Most Popular Social Network Sites (SNS)	YouTube, Facebook, Reddit, Twitter, LinkedIn
10	Large non-Internet companies: Website	JPMorgan Chase, Bank of America, Wells Fargo, AT&T, McLane
	Large non-Internet companies: SNS	Exxon Mobile, Geico, Citigroup, Microsoft
	Large non-Internet companies: Website and SNS	Verizon
38		

Tabelle 3: Ausgewählte Unternehmen aus den Vereinigten Staaten

GB		
3	Web Portal Sites	MSN, Yahoo, AOL
3	Blog Hosting Sites	WordPress, Tumblr, Disqus
3	Community Sites	Momsnet, Stack Exchange, Imgur
3	E-Commerce Sites	Amazon, eBay, Gumtree
3	Recommendation Portals	TripAdvisor, Trustpilot, Rotten Tomatoes
8	Major News Media Sites	The Guardian, Daily Mail, The Telegraph, The Times, The Independent, The Sun, Evening Standard, Financial Times
5	Most Popular Social Network Sites (SNS)	YouTube, Facebook, Twitter, LinkedIn, Reddit
10	Large non-Internet companies: Website	HSBC, British American Tobacco, Diageo, WPP, BAE Systems
	Large non-Internet companies: SNS	Lloyds Banking Group, GlaxoSmithKline, Astra Zeneca
	Large non-Internet companies: Website and SNS	National Grid, Reckitt Benckiser Group
38		

Tabelle 4: Ausgewählte Unternehmen aus Großbritannien

Da viele Unternehmen mehrere unterschiedliche Richtlinien auf ihren Webseiten veröffentlichen, wurden für diese Forschungsarbeit insgesamt 419 Richtlinien von 152 Unternehmen analysiert. Eine Übersicht über Anzahl und Art der analysierten Richtlinien – Terms of Use, Community Guidelines, Content Guidelines und Reporting Guidelines – aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien ist in Tabelle 5 dargestellt.

	Terms of Use	Community Guidelines	Content Guidelines	Reporting Guidelines	Total
AU	32	43	12	16	103
GE	30	44	16	19	109
US	32	31	16	16	95
GB	39	40	15	18	112
Total	133	158	59	69	419

Tabelle 5: Anzahl und Art der analysierten Richtlinien pro Land

### 3.1.3. Kategorienbildung

In dieser Arbeit sollen, wie bereits erwähnt, folgende Forschungsfragen beantwortet werden:

**FF1:** Welche Inhalte zählen laut der Unternehmensrichtlinien zu HOC?

**FF2:** Welche Maßnahmen werden von den Unternehmen getroffen, um gegen HOC vorzugehen?

**FF3:** Inwiefern können länderspezifische Unterschiede in Bezug auf HOC festgestellt werden?

Um die relevanten Kategorien für die detaillierte Inhaltsanalyse festzulegen, wurden zunächst 10% der Richtlinien qualitativ untersucht. Auf Basis der formulierten Forschungsfragen und der qualitativen Voruntersuchung wurde ein Kategorienschema mit fünf Kategorien entwickelt, welches in Tabelle 6 ersichtlich ist.

Kategorie 1: Wie sollen sich User und Userinnen verhalten?
Viele Unternehmen halten in ihren Richtlinien fest, wie sich die User und Userinnen verhalten sollen. Besonders häufig wird beispielsweise die Verantwortlichkeit für die eigenen Inhalte, verantwortungsbewusstes Handeln oder ein respektvoller Umgang erwähnt.
Kategorie 2: Was wird den Userinnen und Usern untersagt?
In ihren Richtlinien kommunizieren die Unternehmen außerdem welche Inhalte und/oder Verhaltensweisen untersagt sind. Darunter fallen unter anderem Pornographie, Hate Speech und Diskriminierungen.
Kategorie 3: Warum dürfen Userinnen und User gewisse Inhalte nicht posten?
In vielen Richtlinien wird begründet, warum gewisse Inhalte nicht gepostet werden dürfen und welches Ziel mit einem Verbot verfolgt wird. Bei diversen Foren gibt es beispielsweise Spielregeln, die beachtet werden müssen, damit Qualität und Niveau gewährleistet bleiben.
Kategorie 4: Was sind die Konsequenzen bei einem Verstoß gegen die Richtlinien?
Bei einem Verstoß gegen die Unternehmensrichtlinien drohen den Userinnen und Usern unterschiedliche Konsequenzen. Viele Unternehmen bevorzugen dabei das Löschen der Kommentare, ohne eine Erklärung abzugeben.
Kategorie 5: Gibt es Anleitungen zur Meldung eines Kommentars? Was passiert mit den gemeldeten Inhalten?
Einigen Unternehmen ist es besonders wichtig, ihren Userinnen und Usern genauestens zu erklären, wie sie gegen die Richtlinien verstoßende Kommentare melden können. Im Anschluss einer Meldung wird oftmals auch berichtet, welche Entscheidung getroffen wurde und was mit dem gemeldeten Inhalt passiert.

Tabelle 6: Kategorien 1-5

#### 5.1.4. Codebuch

Das für diese Forschungsarbeit verwendete Codebuch, das im Anhang ersichtlich ist, basiert auf dem englischsprachigen Codebuch von Einwiller und Kim und wurde um

weitere 16 Variablen ergänzt. Das von Einwiller und Kim erstellte Codebuch enthält 42 Variablen und kann in drei Abschnitte eingeteilt werden.

Der erste Abschnitt des Codebuchs enthält Variablen zur Erfassung grundlegender Informationen über das Dokument (z.B. Name des Unternehmens, Art des Dokuments). Wie bereits erwähnt, wurden sowohl Nutzungsbedingungen als auch Community-, Content-, und Reporting Guidelines untersucht. Die Unternehmen wurden im Codebuch in folgende Kategorien eingeteilt:

- Web Portal Sites (z.B. GMX, Web.de)
- Blog Hosting Sites (z.B. Jimdo, WordPress)
- News Media Sites (z.B. Der Standard, The New York Times)
- Social Network Sites (z.B. Facebook, LinkedIn)
- Community Sites (z.B. Momsnet, Imgur)
- E-Commerce Sites (z.B. Amazon, eBay)
- Corporations' SNSs (z.B. Citigroup, BMW)
- Corporations' Websites (z.B. OMV Group, Red Bull)
- Recommendation Portals (z.B. TripAdvisor, Rotten Tomatoes)

Der zweite Abschnitt konzentriert sich auf die Zugänglichkeit und Lesbarkeit der Richtlinien. Hier wurde unter anderem erhoben, ob eine Registrierung für das Einsehen der Richtlinien notwendig ist, ob die Richtlinie auf der Startseite der Homepage zu finden ist oder wie viele Klicks benötigt werden, um das Dokument zu finden. Außerdem wurde ermittelt, wie leicht navigierbar und leserlich die Dokumente sind, also ob die Richtlinien sprachliche oder bildhafte Beispiele zur Unterstützung der Anschaulichkeit beinhalten.

Im dritten und letzten Abschnitt wurden die Richtlinien inhaltlich untersucht. Hier ging es unter anderem um die verwendeten Formulierungen (fordernde bzw. gleichermaßen verbietende und anleitende Sprache, anleitende oder ausschließlich verbietende Sprache). Es wurde aber auch erfasst, wie Unternehmen mit HOC umgehen bzw. wie sie auf HOC reagieren (z.B. Löschen von Beiträgen, Mahnung, Löschen des Accounts). Außerdem wurde untersucht, welche Handlungsmöglichkeiten die Userinnen und User bei einem Verstoß gegen die Richtlinien haben. Viele Unternehmen geben den Userinnen und Usern nämlich die Möglichkeit, selbständig gegen HOC vorzugehen. Dabei können diese unter anderem E-Mails schreiben, sich telefonisch oder per Post mit den Unternehmen in Verbindung setzen.

Einige der im Codebuch von Einwiller und Kim vorhandenen Variablen waren für diese Forschungsarbeit weniger relevant und wurden daher nicht verwendet. Diejenigen Variablen, die für diese Arbeit von Bedeutung sind, wurden beibehalten. Dabei handelt es sich primär um die bereits erwähnten Variablen zur Erfassung grundlegender Informationen über die Richtlinien (z.B. COMPANY, MEDIUM, COUNTRY) und um einige inhaltliche Variablen (z.B. INTERACT, ACTION 1–8).

Das bestehende Codebuch wurde jedoch auch um weitere Variablen ergänzt. Die Variablen „HOW 1–4“ sollen Antworten auf die Frage liefern, wie – *how* – sich Userinnen und User beim Schreiben von Posts/Kommentaren verhalten sollen. Mit diesen Variablen wurde erhoben, ob in den Richtlinien erwähnt wird, dass Userinnen und User für ihren eigenen Inhalt verantwortlich sind (HOW 1), dass sie sich verantwortungsbewusst verhalten sollen (HOW 2), dass sie gebeten werden, sich gegenseitig zu respektieren (HOW 3) oder dass sie explizit aufgefordert werden, Inhalte zu melden, die gegen die Richtlinie verstoßen (HOW 4).

Außerdem wurde in dem Codebuch erfasst, welche Inhalte und/oder Verhaltensweisen in den Unternehmensrichtlinien untersagt – *forbidden* – werden. Erhoben wurden konkrete (und vergleichbare) Begriffe wie Pornographie (FORBIDDEN 1), Hate Speech (FORBIDDEN 2), Illegal (FORBIDDEN 3), Belästigung (FORBIDDEN 4), Diskriminierung (FORBIDDEN 5), Verletzung der Privatsphäre und Rechte Dritter (FORBIDDEN 6) und Verstöße gegen Kinder- und Jugendschutz (CHILD).

Da viele Unternehmen in ihren Richtlinien auch begründen, wieso sie gewisse Inhalte und/oder Verhaltensweisen verbieten, wurden die „PURPOSE“-Variablen erstellt. Als mögliche Gründe wurden hier die Meinungsfreiheit (PURPOSE 1), die Förderung gesunder Debatten (PURPOSE 2) und die Einhaltung von Verhaltensregeln (PURPOSE 3) miteinbezogen.

Um herauszufinden, welche Konsequenzen den Userinnen und Usern bei einem Verstoß gegen die Unternehmensrichtlinien drohen, mussten keine neuen Variablen erstellt werden, da hierfür einige der bereits im Codebuch von Einwiller und Kim vorhandenen „HANDLE“-Variablen<sup>1</sup> herangezogen werden konnten. Folgende denkbare

---

<sup>1</sup> Hinweis: Im originalen Codebuch wurden elf „HANDLE“-Variablen erstellt. HANDLE 6–9 sind für diese Forschungsarbeit allerdings nicht relevant, da hier nur Richtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien untersucht werden und die besagten Variablen sich auf Konsequenzen aus Südkorea, China oder Japan beziehen.

Konsequenzen wurden dabei erhoben: Löschung ohne Angabe von Gründen (HANDLE 1), Löschung/Einschränkung mit Erklärung bzw. Kommentar (HANDLE 2), Warnung der Userin/des Users (HANDLE 3), Löschung oder Schließung des Benutzerkontos (HANDLE 4), Löschung oder Schließung der gesamten Diskussion (HANDLE 5), rechtliche und gerichtliche Strafverfolgung (HANDLE 10) oder sonstige (HANDLE 11).

Dass Userinnen und User von Unternehmen aufgefordert werden, bei Verstößen gegen ihre Richtlinien beispielsweise Beiträge zu markieren oder den Provider zu benachrichtigen, haben Einwiller und Kim bereits in ihrer Forschungsarbeit dargelegt. In dieser Arbeit wurde noch zusätzlich der Frage nachgegangen, ob in den Richtlinien konkrete Anleitungen zur Meldung von Kommentaren vorhanden sind (INSTRUCTION) und ob erwähnt wird, was mit den gemeldeten Inhalten passiert (REPORTED).

#### 5.1.5. Pretest

Ziel einer Probecodierung ist es, die Genauigkeit des Codebuchs zu überprüfen. Hierfür wurden zunächst 10% der Unternehmensrichtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien (N = 42), welche in Tabelle 7 ersichtlich sind, codiert. Es wurden bewusst unterschiedliche Richtlinien (Nutzungsbedingungen, Community Guidelines, Content Guidelines sowie Reporting Guidelines) codiert, um herauszufinden, ob die bestehenden Kategorien adaptiert oder erweitert werden müssen.

	Anzahl der Dokumente	10% der Dokumente	Codierte Richtlinien
AU	103	10	AUgmx1, AUgmx2-1, AUgmx2-2, AUweb.de1, AUweb.de2-1, AUweb.de2-2, AUyahoo1, AUyahoo2-1, AUyahoo2-2, AUyahoo4
GE	109	11	GEgmx1, GEgmx2-1, GEgmx2-2, GEyahoo1, GEyahoo2-1, GEyahoo2-2, GEyahoo4, GEweb.de1, GEweb.de2-1, GEEbay3, GEallianz2
US	95	10	USaol1, USaol2, USweebly1, USweebly4, USwordpress1, USwordpress2, USblogger3, USusatoday1, USstackexchange3, USyelp3
GB	112	11	UKtumblr2, UKdisqus1-1, UKdisqus1-2, UKdisqus2-1, UKdisqus2-2, UKdisqus2-3, UKdisqus2-4, UKdisqus3-1, UKdisqus3-2, UKthesun1, UKthetimes2 <sup>2</sup>

Tabelle 7: Anzahl der codierten Unternehmensrichtlinien für den Pretest

Die bestehenden Kategorien hielten der Probecodierung des Pretests stand und wurden daher nahezu unverändert übernommen. Lediglich einige Codieranweisungen wurden präziser definiert. Um Missverständnisse bei der Codierung zu vermeiden, wurden außerdem auch einige Ankerbeispiele aus dem Textmaterial (den Unternehmensrichtlinien) eingefügt.

### 5.1.6. Reliabilitätstest

Damit eine Inhaltsanalyse von einem anderen Codierer fehlerfrei wiederholt werden kann, muss das Messinstrument, in dem Fall das Codebuch, zuverlässig sein (Brosius, Koschel, & Haas, 2008, S. 152). Um die Zuverlässigkeit bzw. die Reliabilität zu messen, wurde das Codebuch an 20% aller verwendeten Richtlinien aus Österreich, Deutschland, den

<sup>2</sup> Hinweis: Die codierten Richtlinien aus Großbritannien wurden mit „UK“ beschriftet, gemeint ist allerdings „GB“.



Vereinigten Staaten und Großbritannien (N = 84) getestet. Die für den Reliabilitätstest ausgewählten Unternehmensrichtlinien sind in der folgenden Tabelle 8 dargestellt.

	Anzahl der Dokumente	20% der Dokumente	Codierte Richtlinien
AU	103	21	AUgmx1, AUgmx2-1, AUgmx2-2, AUweb.de1, AUweb.de2-1, AUweb.de2-2, AUyahoo1, AUyahoo2-1, AUyahoo2-2, AUyahoo4, AUtumblr1, AUtumblr2, AUtumblr4, AUjimdo1, AUwordpress1, AUwordpress2, AUoe241, AUoe242, AUkrone1, AUkrone2, AUdiepresse2
GE	109	22	GEgmx1, GEgmx2-1, GEgmx2-2, GEyahoo1, GEyahoo2-1, GEyahoo2-2, GEyahoo4, GEweb.de1, GEweb.de2-1, GEweb.de2-2, GETumblr1, GETumblr2, GETumblr4, GEwordpress1, GEwordpress2, GEblogger3, GESüddeutsche1, GEfocus1, GEfocus2, GESpiegel1, GESpiegel2, GEzeit1
US	95	19	USaol1, USaol2, USyahoo1, USyahoo2-1, USyahoo2-2, USyahoo4, USweebly1, USweebly4, USwordpress1, USwordpress2, USblogger3, USusatoday1, USusatoday2-1, USusatoday2-2, USnydailynews1, USnydailynews2, USwashingtonpost2, USchigacotribune1, USlatimes1
GB	112	22	UKaol1, UKaol2, UKyahoo2-2, UKyahoo4, UKtumblr1, UKtumblr2, UKwordpress1, UKwordpress2, UKdisqus1-1, UKdisqus1-2, UKdisqus2-1, UKdisqus2-2, UKdisqus2-3, UKdisqus2-4, UKdisqus3-1, UKdisqus3-2, UKdisqus4-1, UKdisqus4-2, UKtheindependent1, UKtheindependent2, UKthetelegraph1, UKthetelegraph2

Tabelle 8: Anzahl der codierten Unternehmensrichtlinien für den Reliabilitätstest

Um die Intracoderreliabilität messen zu können, wurden dieselben Richtlinien sechs Wochen nach dem ersten Codierdurchgang erneut codiert. Die Zuverlässigkeit ist in diesem Fall durch den zeitlichen Abstand zwischen den Codierungen gegeben, da davon ausgegangen werden kann, dass beim zweiten Codierdurchgang nicht aus dem Gedächtnis codiert wurde (Brosius, Koschel, & Haas, 2008).

Im Anschluss der beiden Codierungsdurchgänge wurde ein Reliabilitätstest mittels ReCal („Reliability Calculator“) durchgeführt. Dabei wurde die Übereinstimmung der Codierungen zwischen dem ersten und dem zweiten Codierdurchgang und somit die Zuverlässigkeit des Codierers gemessen (Brosius, Koschel, & Haas, 2008). Angestrebt wurden Koeffizienten von  $\geq .90$ , weil diese Werte laut Neuendorf (2002) in der Methodik als zuverlässig angesehen werden. Das Codebuch bzw. die Codieranweisungen waren im zuvor durchgeführten Pretest bereits konkretisiert und mit Ankerbeispielen versehen worden. Dadurch fiel der Reliabilitätstest mit einer hundertprozentigen Übereinstimmung überdurchschnittlich positiv aus. Die Ergebnisse des ReCal-Reliabilitätstests, welche im Anhang ersichtlich sind, haben somit ergeben, dass das Messinstrument zuverlässig ist und dass die Codierungen sorgfältig und korrekt vorgenommen wurden.

Um die Zuverlässigkeit der Ergebnisse der empirischen Untersuchung zu garantieren, wurde während der laufenden Codierung ein weiterer Reliabilitätstest durchgeführt. Dieser hat den Anforderungen ebenfalls standgehalten.

## 5.2. Ergebnisse der Inhaltsanalyse

In diesem Kapitel werden die wichtigsten Erkenntnisse der Inhaltsanalyse vorgestellt und die einzelnen Forschungsfragen beantwortet. Zunächst werden allerdings die formalen Merkmale der Unternehmensrichtlinien dargestellt.

Wie bereits bei der Vorstellung der Methode erwähnt, wurden insgesamt 419 Richtlinien von 152 Unternehmen analysiert. Darunter 103 Richtlinien aus Österreich (24,6%), 109 aus Deutschland (26%), 95 aus den Vereinigten Staaten (22,7%) sowie 112, also 26,7% aus Großbritannien (siehe Abbildung 3).

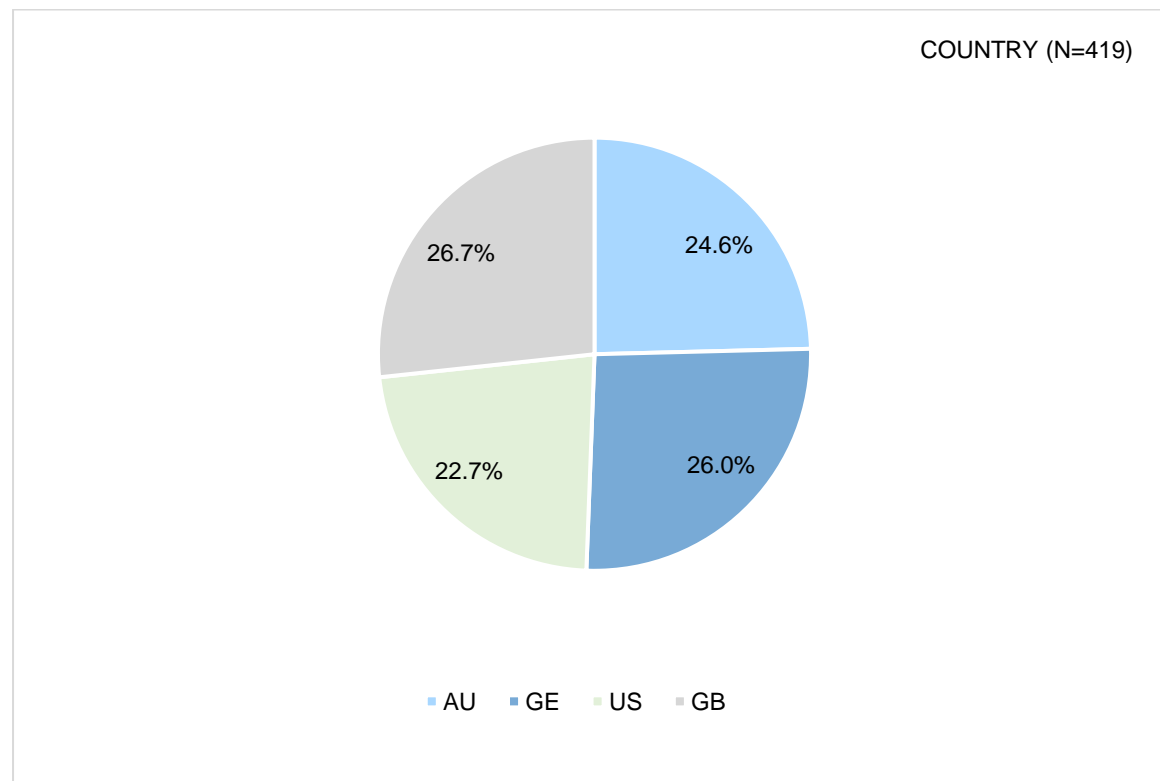


Abbildung 3: Herkunft der untersuchten Richtlinien (in %)

Wie in Abbildung 4 zu sehen ist, wurden am häufigsten Social Network Sites, also Richtlinien von Unternehmen wie Facebook, Instagram, YouTube oder LinkedIn, untersucht. 56 der 419 analysierten Richtlinien stammten von Medienunternehmen, 35 von Blog Hosting Sites wie WordPress oder Tumblr und 33 von Web Portal Sites wie

Yahoo oder MSN. Außerdem wurden 28 Richtlinien von Corporations' SNSs und jeweils 23 Richtlinien von Community Sites (z.B. Momsnet, Imgur) und von E-Commerce Sites (z.B. Amazon, eBay) analysiert. Lediglich 22 der 419 Richtlinien stammten von Recommendation Portals wie beispielsweise TripAdvisor und 21 von Corporations' Websites.

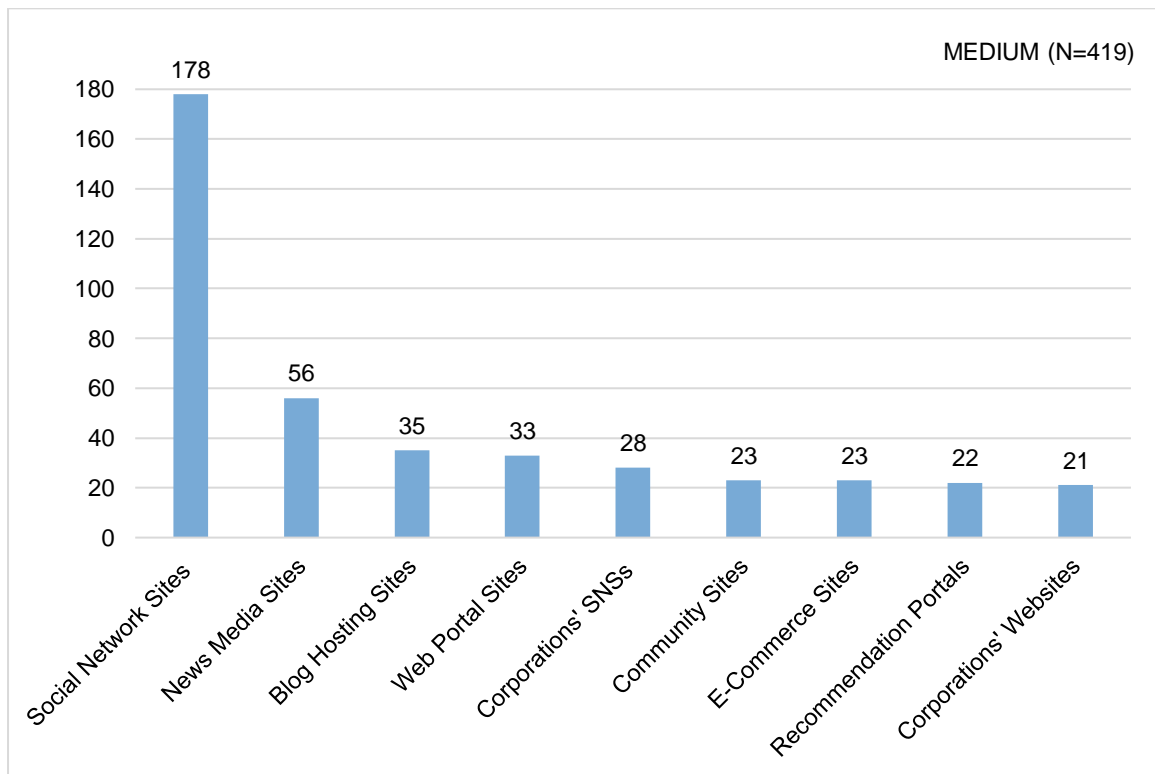


Abbildung 4: Anzahl der Dokumente aus den jeweiligen Kategorien

Die Richtlinien wurden im Codebuch nach ihrer Art in folgende Kategorien eingeteilt: (1) Terms of Use (darunter auch Terms of Service, Conditions of Use, Service Agreements oder Operation Policies); (2) Community Guidelines (darunter auch Community Standards, Netiquetten, Missions, Rules); (3) Content Guidelines und (4) Reporting Guidelines. Insgesamt 37,7% der analysierten Richtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien waren Community Guidelines. 31,7% der Richtlinien waren Terms of Use, 16,5% Reporting Guidelines und 14,1% Content Guidelines (siehe Abbildung 5).

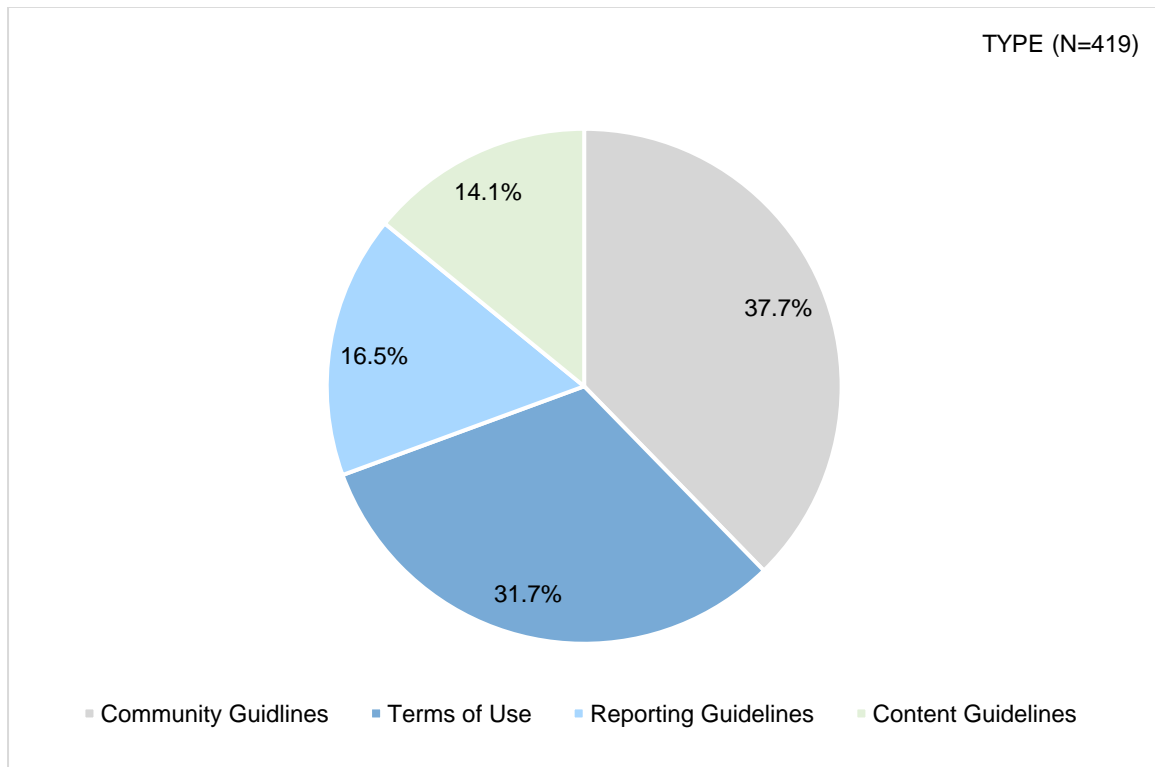


Abbildung 5: Anteil der analysierten Richtlinien nach Art (in %)

### 5.2.1. Inhalte der Richtlinien

Ziel dieser Forschungsarbeit ist es unter anderem herauszufinden, welche Inhalte laut der Unternehmensrichtlinien zu HOC zählen. Hierfür wurde untersucht, welche Inhalte und/oder Verhaltensweisen den Userinnen und Usern untersagt werden. Wie in Abbildung 6 zu sehen ist, wurde in 80,2% der 419 Richtlinien kommuniziert, dass Harassment also etwa Belästigungen, Mobbing, Verleumdungen oder andere abwertende, beleidigende, einschüchternde oder bedrohende Äußerungen bzw. Inhalte nicht geduldet werden. In 70,6% der Dokumente kam vor, dass illegale und gewalttätige Inhalte bzw. kriminelles Verhalten verboten ist. Pornographische oder sonstige obszöne Inhalte zählen in 52,3% der Richtlinien zu HOC. Die Diskriminierung von Personen aufgrund des Geschlechtes, der Herkunft, der Religion, der Hautfarbe oder der sexuellen Orientierung wird in 41,5% der Richtlinien nicht toleriert. Hate Speech und sonstige hasserfüllte Inhalte werden hingegen nur in 35,6% der Unternehmensrichtlinien verboten. Nur das Verbot von Datenschutzverletzungen und die Verletzung von Rechten Dritter werden mit 32% seltener in den Richtlinien erwähnt. In 107 von 419 Richtlinien (25,5%), wurde außerdem

explizit erwähnt, dass gewisse Inhalte aus Kinder- und Jugendschutzgründen auf den Plattformen verboten werden.

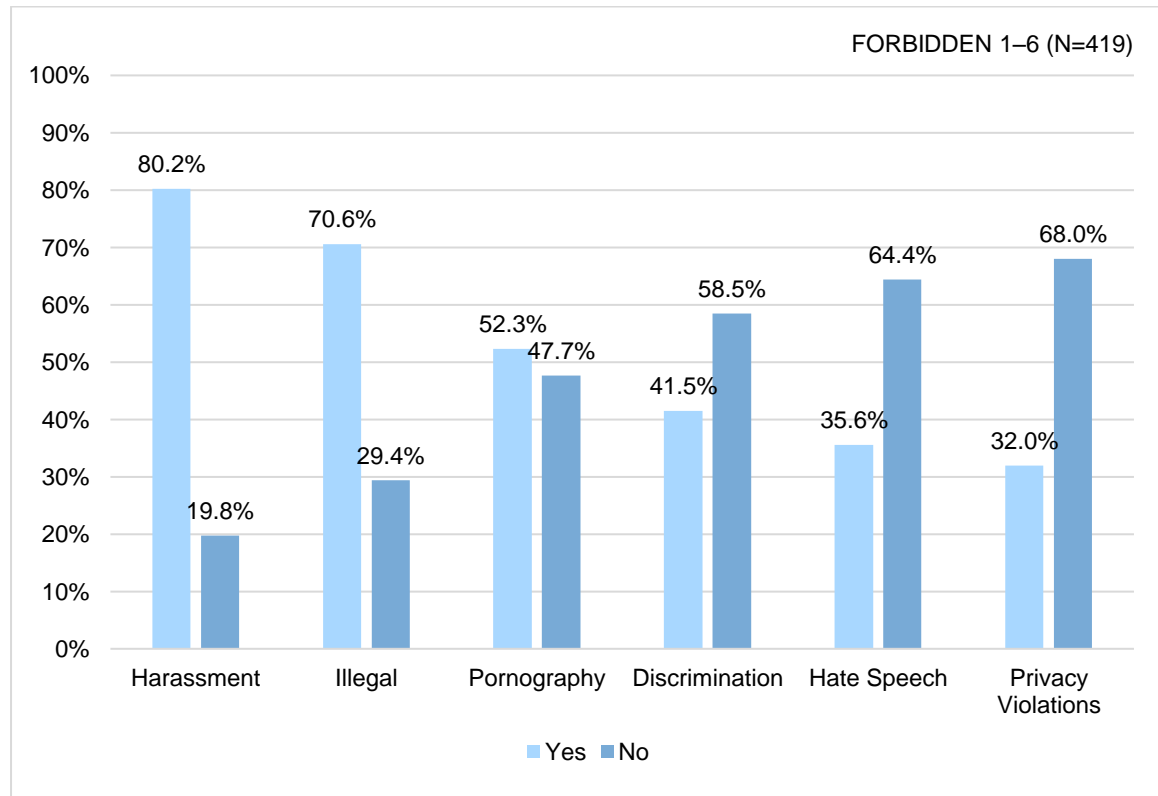


Abbildung 6: Einstufung der unterschiedlichen Inhalte als HOC laut den Richtlinien

## 5.2.2. Maßnahmen gegen HOC

Diese Arbeit befasst sich außerdem mit den Maßnahmen, die von den Unternehmen getroffen werden, um gegen HOC vorzugehen. Eine wichtige Maßnahme ist das Erstellen von Verhaltensrichtlinien, in denen die Unternehmen unter anderen auch kommunizieren können, wie sich die Userinnen und User auf ihren Plattformen verhalten sollen. In 148 der 419 untersuchten Richtlinien (35,3%) wurden Beispiele angeführt, wie die Userinnen und User miteinander umgehen bzw. interagieren sollen (siehe Abbildung 7).

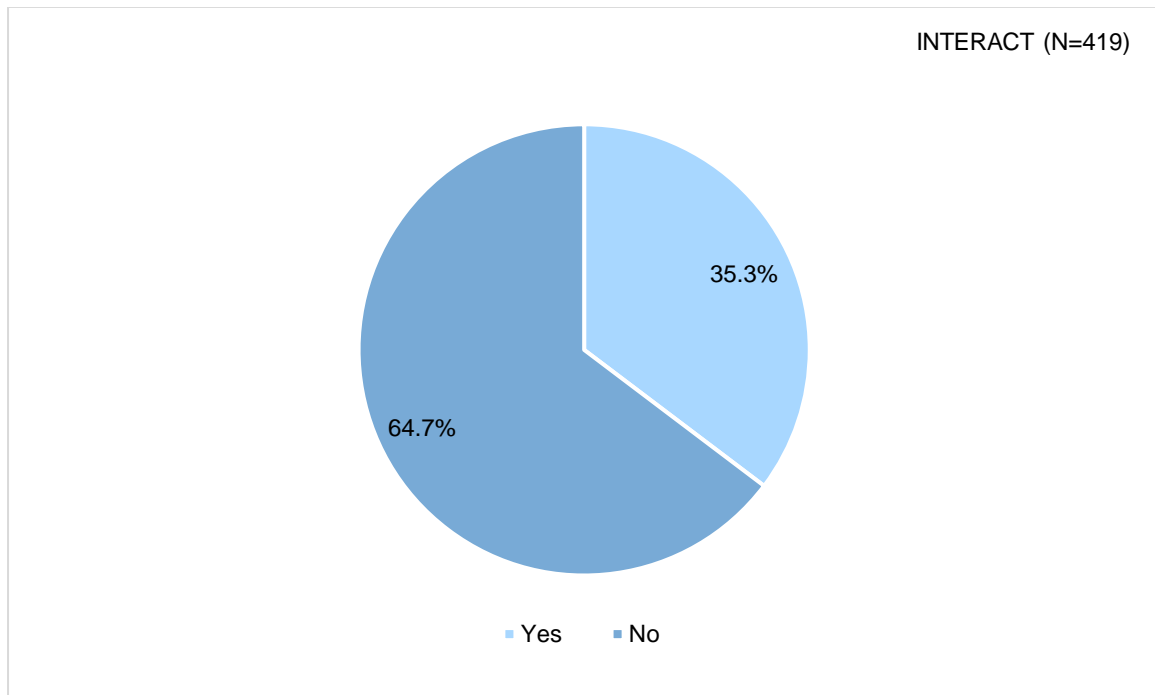


Abbildung 7: Anteil der Richtlinien, in denen erwähnt wird, wie Userinnen und User interagieren sollen (in %)

Wie in Abbildung 8 zu sehen ist, werden aber auch andere Verhaltensweisen erwünscht. In 31% der Richtlinien wurden Userinnen und User ausdrücklich aufgefordert, Inhalte zu melden, die gegen die Unternehmensrichtlinien verstoßen. In 24,3% der Richtlinien wurden die Userinnen und User darauf aufmerksam gemacht, dass sie für ihre eigenen Inhalte verantwortlich sind und in 17,4% der Richtlinien wurden sie gebeten, einander mit Respekt zu begegnen. In 5,5% aller Richtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien wurde von den Userinnen und Usern ein verantwortungsbewusstes Verhalten verlangt.

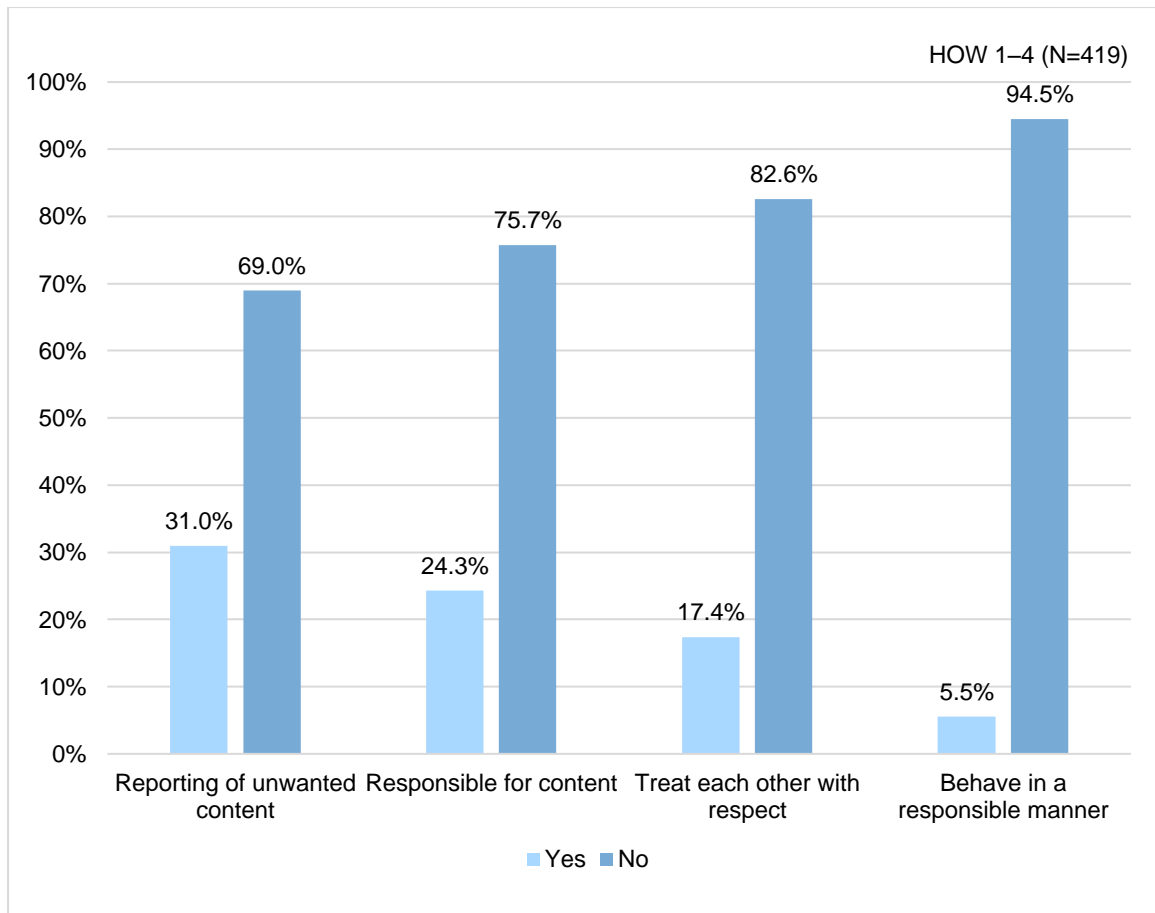


Abbildung 8: Erwünschtes Verhalten, das in den Richtlinien von den Userinnen und Usern erwartet wird

Das Melden von Inhalten, die gegen Unternehmensrichtlinien verstoßen, ist eine wichtige Maßnahme, um gegen HOC vorzugehen. Hierfür bieten die Unternehmen auch unterschiedliche (technische) Möglichkeiten an. In den meisten Fällen haben die Userinnen und User die Möglichkeit, regelwidrige Inhalte zu markieren und dadurch direkt zu melden. Aus Abbildung 9 geht hervor, dass diese Option in 20,3% der Fälle explizit in den Richtlinien erwähnt wurde. Darüber hinaus gab es auch in 43,9% der Fälle die Möglichkeit, Inhalte (auch ohne explizite Erwähnung in der Richtlinie) direkt auf der Plattform zu melden. Viele Unternehmen entwickeln auch ein eigenes elektronisches Meldesystem, um den Userinnen und Usern den Vorgang des Meldens, beispielsweise mit Hilfe von Anweisungen, zu vereinfachen. Von den insgesamt 419 analysierten Richtlinien wurde diese Möglichkeit in 36% der Fälle ohne explizite Erwähnung angeboten. In 22,2% der Richtlinien wurde auch ausdrücklich auf diese Option hingewiesen. Darüber hinaus bieten einige Plattform-Betreiber auch die Möglichkeit, Verstöße via Email, Post oder Telefon zu melden. Diese Möglichkeiten wurden allerdings vergleichsweise eher selten angeboten. Interessant ist auch, dass in lediglich 4,3% der Richtlinien erwähnt



wurde, dass Userinnen und User in Form von Counter Speech gegen HOC vorgehen sollten. Noch seltener, nämlich in 3 von 419 Richtlinien (0,7%) raten die Unternehmen, Verstöße über einen angeführten Link an entsprechende Behörden zu melden. Abgesehen von den bisher erwähnten Handlungsmöglichkeiten für Userinnen und User, werden in 14,1% der Richtlinien auch andere Optionen genannt. Zu den alternativen Handlungsmöglichkeiten zählen unter anderem das Blockieren und das temporäre oder dauerhafte Sperren von Userinnen und Usern. Personen erhalten auch auf vielen Plattformen die Möglichkeit, zu entscheiden, welche Inhalte sie sehen und welche sie ausblenden wollen.

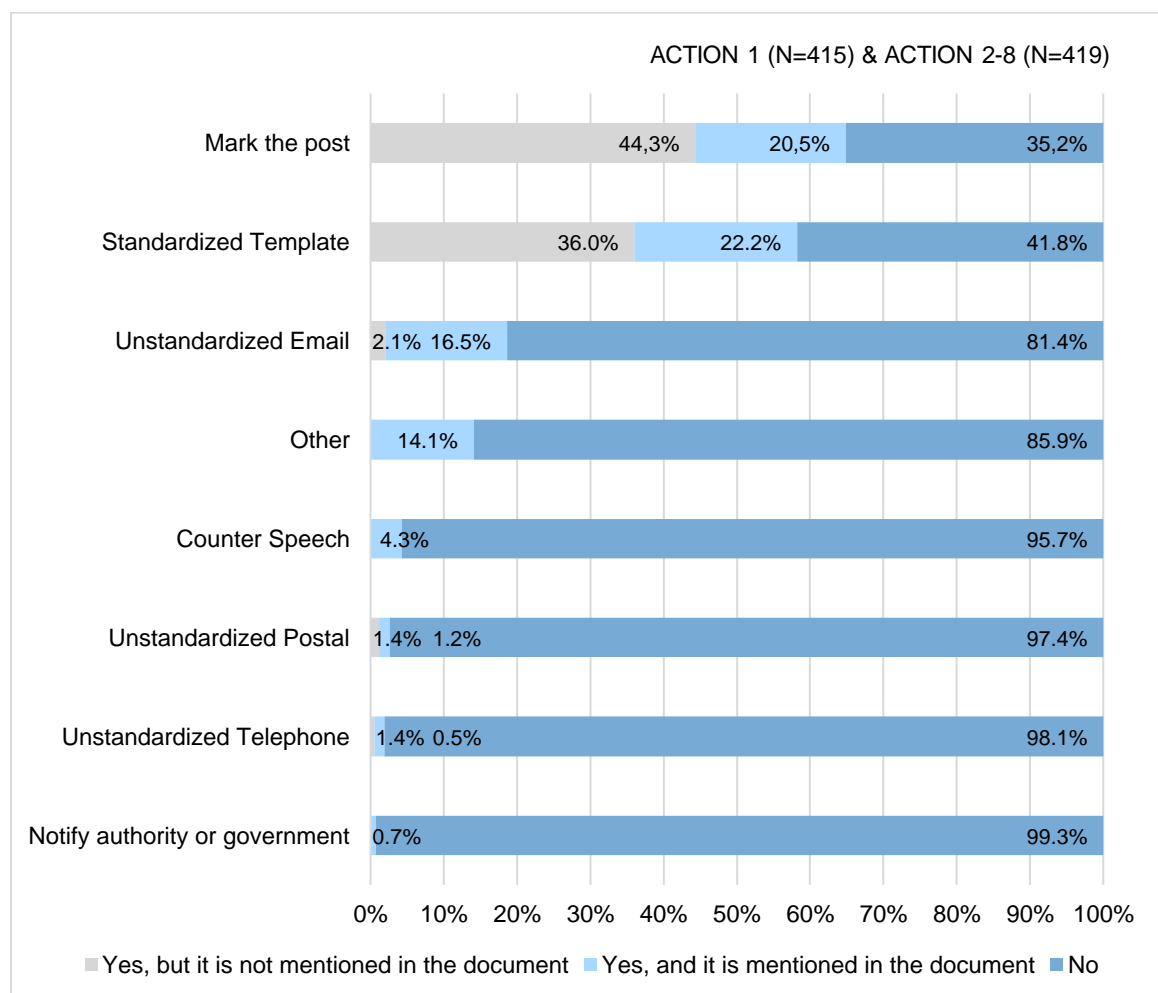


Abbildung 9: Handlungsmöglichkeiten der Userinnen und User in Bezug auf HOC

Um effektiv gegen HOC vorgehen zu können, sind die Unternehmen unter anderem auf die Userinnen und User ihrer Plattformen angewiesen. Sie werden gebeten, gegen die

Richtlinien verstoßende Inhalte zu melden. Wie soeben dargelegt wurde, werden den Userinnen und Usern hierfür unterschiedliche Möglichkeiten angeboten. Um den Überblick über die diversen Optionen nicht zu verlieren bzw. um eine Meldung auch korrekt durchzuführen, enthalten 22,4% der Richtlinien eine detaillierte Anleitung für die Meldung von regelwidrigen Inhalten (siehe Abbildung 10).

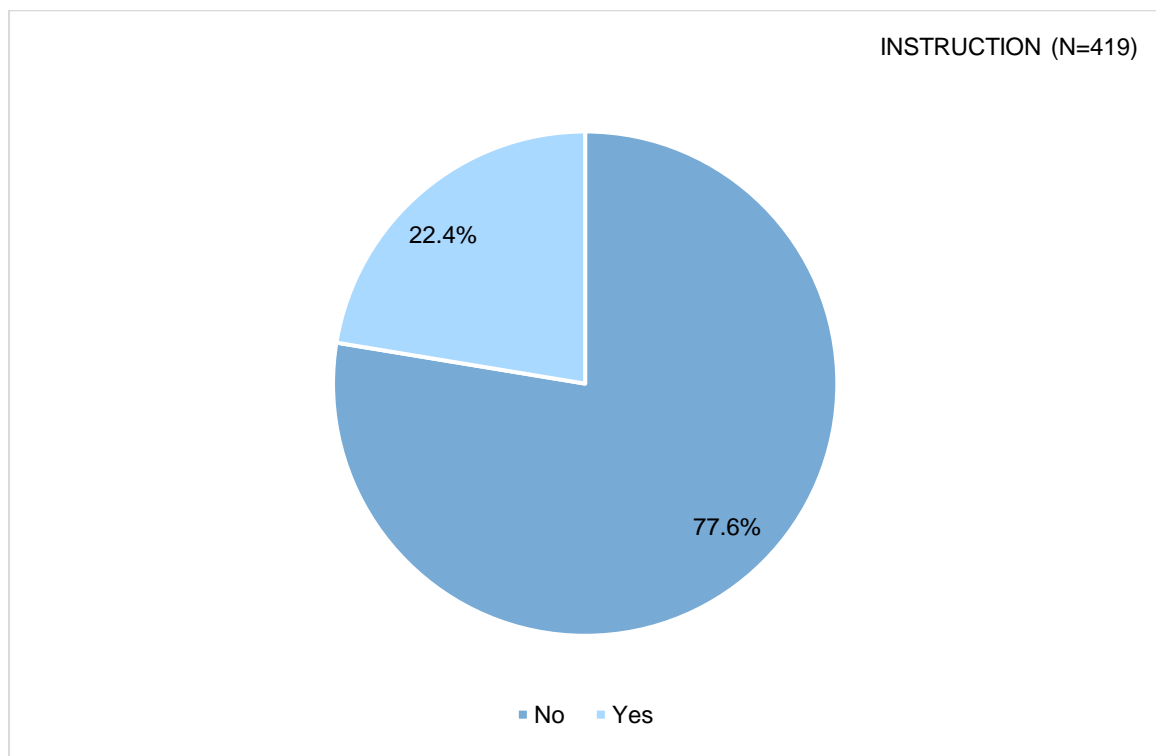


Abbildung 10: Anteil der Richtlinien, die eine detaillierte Anleitung zum Melden von HOC beinhalten (in %)

Aus Abbildung 11 geht hervor, dass in 18,1% der 419 Richtlinien explizit erwähnt wird, welche Schritte von den Unternehmen eingeleitet werden, um die gemeldeten Inhalte zu überprüfen. Userinnen und User, die regelwidrige Inhalte an die Plattform-Betreiber melden, erfahren dadurch was mit ihren gemeldeten Inhalten passiert.

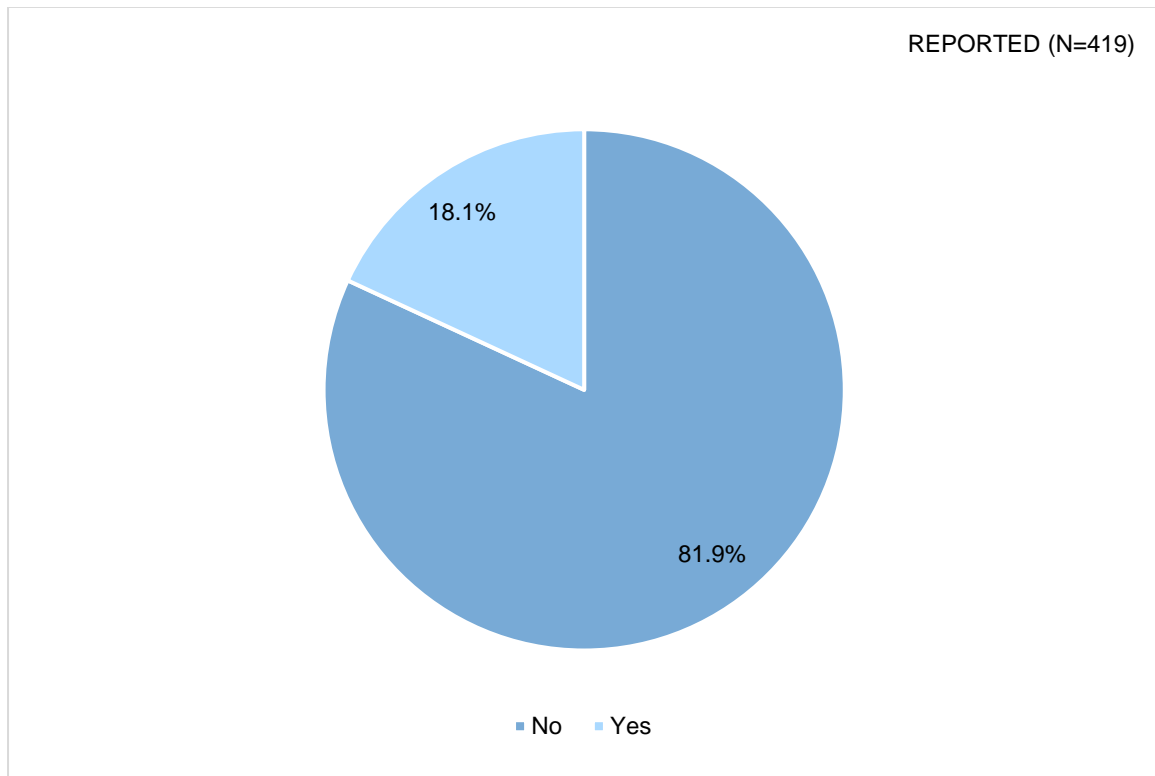


Abbildung 11: Anteil der Richtlinien in denen erwähnt wird, wie die Unternehmen mit den gemeldeten Inhalten umgehen (in %)

Viele Unternehmen informieren ihre Userinnen und User darüber, weshalb gewisse Inhalte oder Verhaltensweisen auf ihren Plattformen verboten sind. Diese Maßnahme der Unternehmen soll einerseits für mehr Transparenz und andererseits für einen respektvollen Umgang auf den Online-Plattformen sorgen. Die Gründe, die laut den Unternehmen für ein Verbot bestimmter Inhalte und Verhaltensweisen sprechen, werden in Abbildung 12 angeführt. In 12,4% der untersuchten Richtlinien erwähnen Plattform-Betreiber, dass sie dazu verpflichtet sind, trotz des hohen Stellenwerts der freien Meinungsäußerung, gewisse Inhalte oder Verhaltensweisen zu unterbinden. Ein weiterer Grund, der ein Verbot von bestimmten Inhalten rechtfertigen soll, ist die Förderung einer konstruktiven Debatte bzw. eines gesunden Diskussionsklimas. In 8,1% der Dokumente werden deshalb auch die Userinnen und User angeregt, objektive und faire Diskussionen zu führen. Außerdem plädieren 8,1% der Unternehmen in ihren Richtlinien für ein verantwortungsbewusstes bzw. ethisches Verhalten auf ihren Plattformen. Die verschriftlichten Verhaltensregeln („rules of good conduct“) sollen dies gewährleisten.

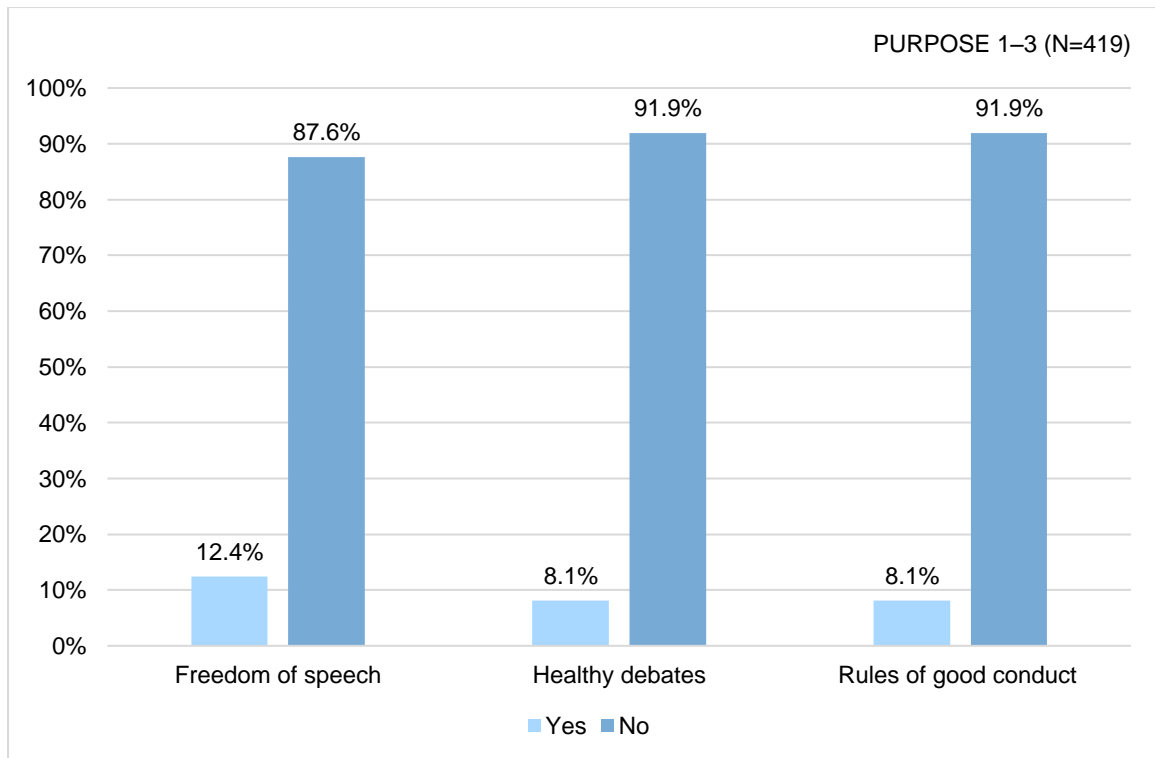


Abbildung 12: Gründe für ein Verbot von HOC

Unternehmen kommunizieren in ihren Richtlinien aber nicht nur, welche Inhalte verboten sind, wie sich die Userinnen und User verhalten sollen bzw. wie sie selbstständig gegen HOC vorgehen können oder warum gewisse Inhalte nicht toleriert werden. In ihren Richtlinien halten die Plattform-Betreiber auch fest, wie sie mit Verstößen umgehen. Für den Fall, dass trotz der Verhaltensregeln gegen die Richtlinien verstoßen wird, drohen den Userinnen und Usern unterschiedliche Konsequenzen (siehe Abbildung 13). Aus den untersuchten Richtlinien geht hervor, dass es mehrere Möglichkeiten gibt, wie Unternehmen mit HOC umgehen. Die am häufigsten getroffene Maßnahme, nämlich in 65,9% der Fälle, ist die Löschung von regelwidrigen Inhalten ohne Angabe von Gründen. In 55,1% der Fälle droht bei einem Verstoß das Löschen oder Sperren des Benutzerkontos. Aus 25,3% der Richtlinien geht hervor, dass Personen, die regelwidrige Inhalte veröffentlichen, verklagt bzw. gerichtlich verfolgt werden. In einigen wenigen Richtlinien (15,3%) werden die Nutzerinnen und Nutzer bei einem Verstoß zunächst verwarnet, bevor weitere Schritte eingeleitet werden. Maßnahmen, die nur ganz selten getroffen werden, sind das Löschen oder Kürzen von Inhalten mit einer Angabe von Gründen (10,7%) oder das Löschen oder Schließen der gesamten Diskussion (4,5%).

Abgesehen von den bisher erwähnten Konsequenzen bei Verstößen gegen die Richtlinien, werden in 20,5% der Fälle auch andere Möglichkeiten genannt, wie mit HOC umgegangen wird. Zu den alternativen Handlungsmöglichkeiten zählen unter anderem das Kürzen von regelwidrigen Inhalten oder das temporäre Sperren von Userinnen und Usern. Bei strafrechtlich relevanten Inhalten behalten sich auch einige Unternehmen das Recht vor, die IP-Adressen der entsprechenden Userinnen und User an Behörden zu übermitteln.

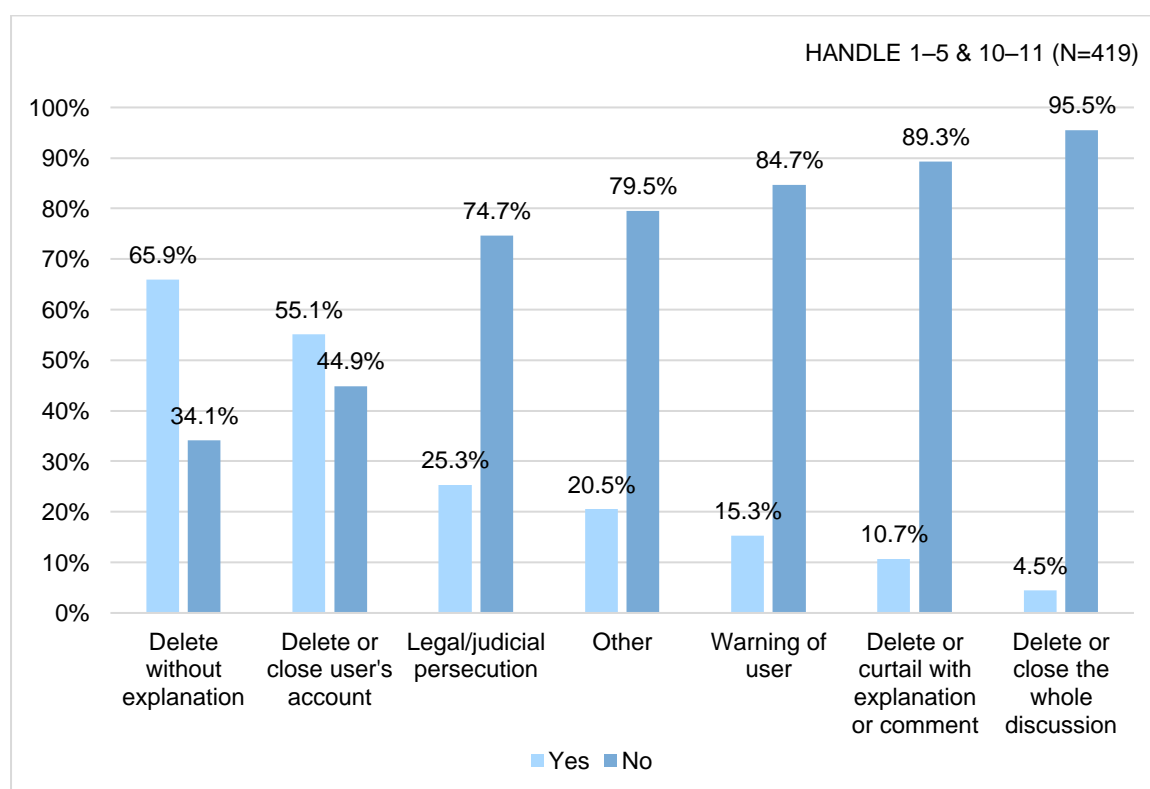


Abbildung 13: Konsequenzen beim Verstoß gegen die Richtlinien

### 5.2.3. Länderspezifische Unterschiede

In dieser Forschungsarbeit wird auch der Frage nachgegangen, ob es länderspezifische Unterschiede im Umgang mit HOC gibt. Zunächst soll festgestellt werden, ob es Unterschiede bei der Art der Richtlinien gibt. Der Anteil der unterschiedlichen Richtlinien, die pro Land untersucht wurden, ist in Abbildung 14 dargestellt. Insgesamt wurden für diese Forschungsarbeit mit 37,7% am häufigsten Community Guidelines untersucht,

gefolgt von Terms of Use (31,7%), Reporting Guidelines (16,5%) und Content Guidelines (14,1%). In den Vereinigten Staaten kommunizieren Unternehmen ihre HOC-Richtlinien mit 32,6% am häufigsten über Terms of Use (N=95). Unternehmen in Österreich, Deutschland und Großbritannien kommunizieren hingegen ihre HOC-Richtlinien hauptsächlich über Community Guidelines (AU: 41,7%, N=103; GE: 40,4%, N=109; GB: 35,7%, N=112). Zwar sind einige Unterschiede zwischen den jeweiligen Ländern hinsichtlich der Art der Richtlinien zu erkennen, dennoch sind diese nicht signifikant:  $X^2(9, N=419) = 3,637, p = 0,934$ . Der Wert bei Cramers V von 0,054 zeigt, dass ein geringer Zusammenhang zwischen den Ländern und der Art der Richtlinien besteht.

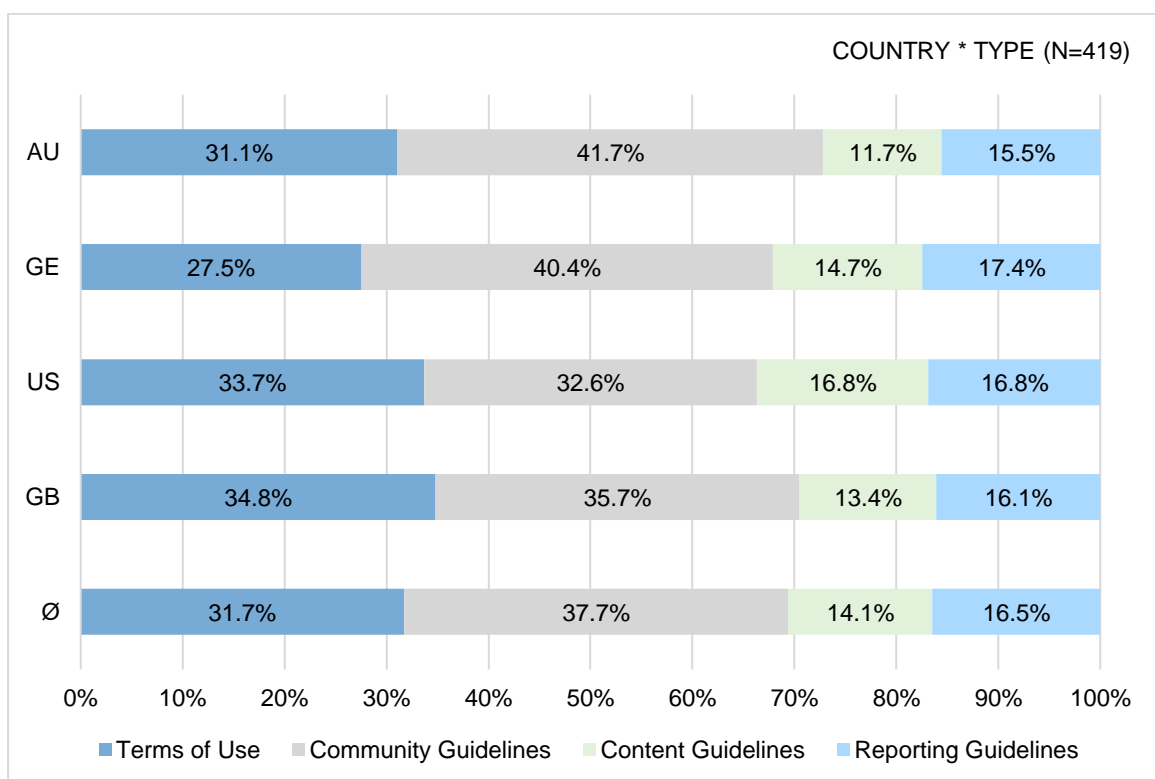


Abbildung 14: Anteil der unterschiedlichen Richtlinienarten nach Land (in %)

In Abbildung 15 ist der Anteil der verbotenen Inhalte pro Land abgebildet. Es fällt auf, dass Hate Speech in österreichischen Richtlinien mit 34,0% genauso oft erwähnt wird wie Datenschutzverletzungen bzw. Verletzungen von Rechten Dritter. Durchschnittlich wird hingegen Hate Speech mit 35,6% öfters erwähnt als die Verletzung von Rechten Dritter mit 32,0%. Aus den Ergebnissen der Vereinigten Staaten geht hervor, dass Hate Speech ein höherer (41,1%) Stellenwert beigemessen wird als Diskriminierungen (34,7%) und

Datenschutzverletzungen (32,6%). In Großbritannien werden, anders als in den anderen Ländern, Datenschutzverletzungen mit 34,8% öfter angesprochen als Hate Speech mit 33,9%. Auch hier zeigt sich, dass die Zusammenhänge gering und die Ergebnisse nicht signifikant sind.

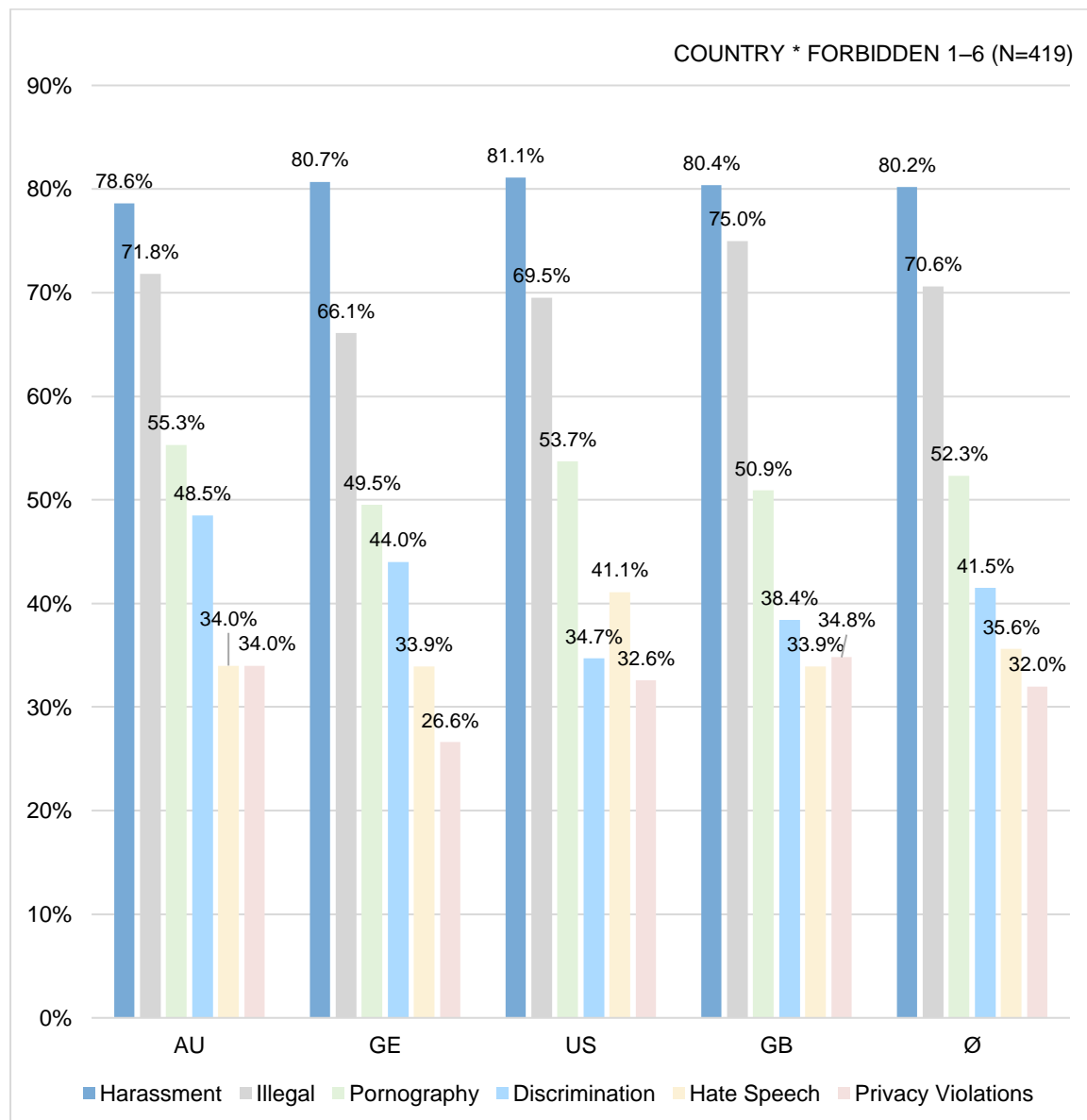


Abbildung 15: Einstufung der unterschiedlichen Inhalte als HOC laut den Richtlinien nach Land (in %)

Es konnte auch kein signifikanter Unterschied zwischen den Ländern hinsichtlich der Erwähnung von Kinder- und Jugendschutzbestimmungen festgestellt werden:  $\chi^2(3, N=419) = 3,315, p = 0,346$ . Die Stärke des Zusammenhangs ist hier mit 0,089 gering.

Aus allen untersuchten Ländern wurden Kinder- und Jugendschutzbestimmungen in Deutschland mit 31,2% am häufigsten in den Richtlinien erwähnt. In Großbritannien kamen mit 20,5% Kinder- und Jugendschutzbestimmungen deutlich seltener vor (siehe Abbildung 16).

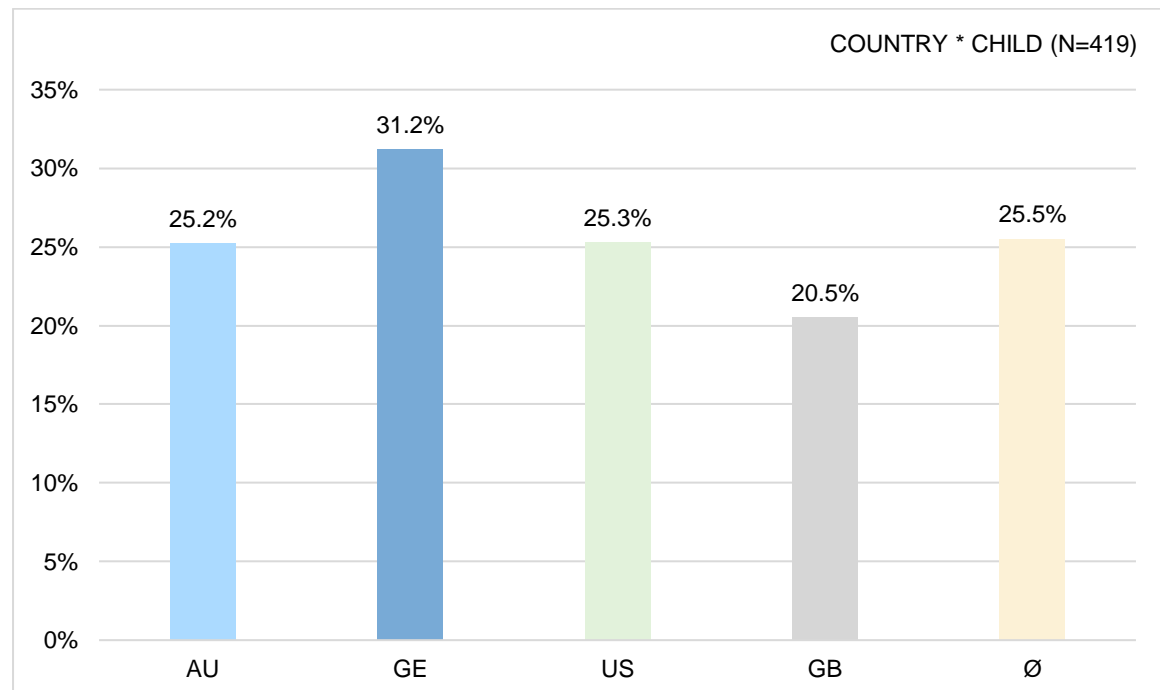


Abbildung 16: Anteil der Erwähnungen von Kinder- und Jugendschutzbestimmungen nach Land (in %)

Ein signifikanter Unterschied konnte hingegen hinsichtlich der Adressierung vom erwünschten Umgang auf den Online-Plattformen festgestellt werden:  $\chi^2(3, N=419) = 11,321, p = 0,010$ . Hier war allerdings die Stärke des Zusammenhangs mit 0,164 gering. Durchschnittlich wurden in 35,5% der Richtlinien Beispiele angeführt, wie Userinnen und User miteinander interagieren sollen (siehe Abbildung 17). Am häufigsten wurde die Interaktion mit 42,2% in deutschen und mit 41,7% in österreichischen Unternehmensrichtlinien erwähnt. In Großbritannien wurde das Thema hingegen mit 23,2% deutlich seltener angesprochen.



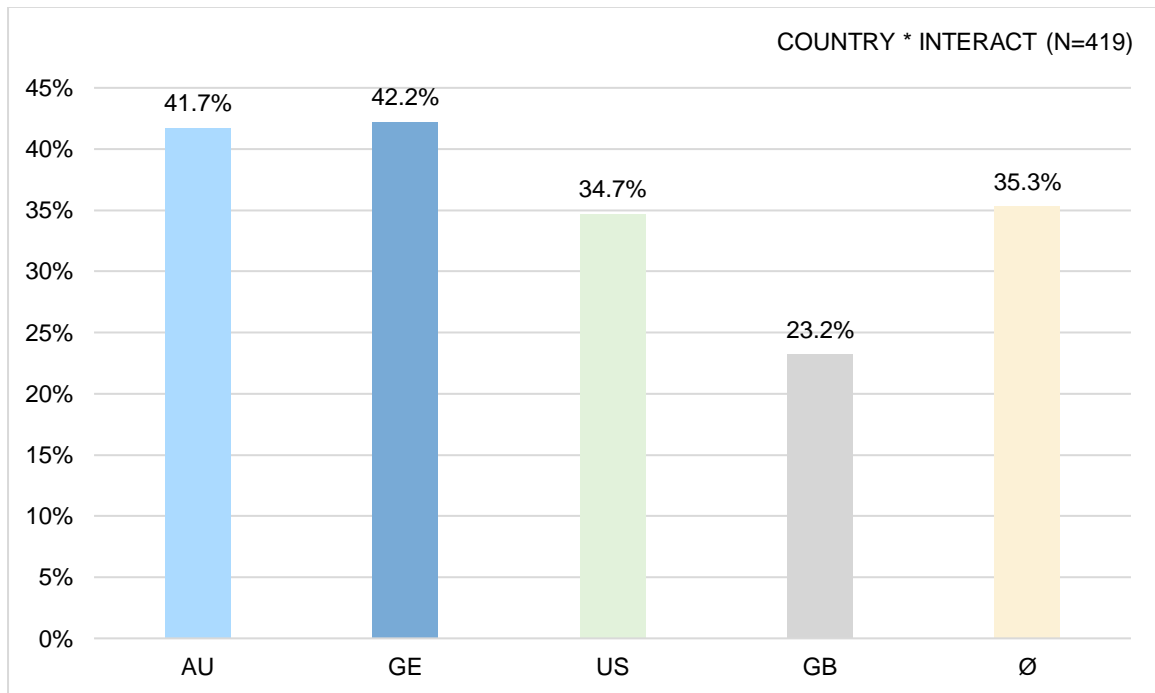


Abbildung 17: Anteil der Richtlinien, in denen erwähnt wird, wie Userinnen und User interagieren sollen, nach Land (in %)

Bei den anderen erwünschten Verhaltensweisen, wie etwa dem Melden von regelwidrigen Inhalten oder dem verantwortungsbewussten Verhalten, konnte hingegen kein signifikanter Unterschied festgestellt werden. Hier unterscheiden sich die Ergebnisse in den jeweiligen Ländern kaum voneinander. Aus Abbildung 18 geht hervor, dass in Österreich das Melden von unerwünschten Inhalten mit 33,0% überdurchschnittlich oft kommuniziert wird. In Deutschland wird es mit 29,4% etwas seltener erwähnt. Ein respektvoller Umgang wird mit 24,3% am häufigsten in den österreichischen Richtlinien gefordert. Deutlich seltener (12,6%) wird das Thema in den Unternehmensrichtlinien aus den Vereinigten Staaten angesprochen. Etwas größere länderspezifische Unterschiede sind außerdem bei der Verantwortlichkeit für die eigenen Inhalte zu beobachten. Diese wird in den Vereinigten Staaten mit 32,6% am häufigsten angesprochen. Einen deutlich geringeren Wert (21,4%) wird ihr hingegen in Großbritannien beigemessen. Verantwortungsbewusstes Verhalten auf den Online-Plattformen wird sowohl in Österreich (7,8%) als auch in Deutschland (5,5%), den Vereinigten Staaten (4,2%) und Großbritannien (4,5%) am seltensten verlangt.

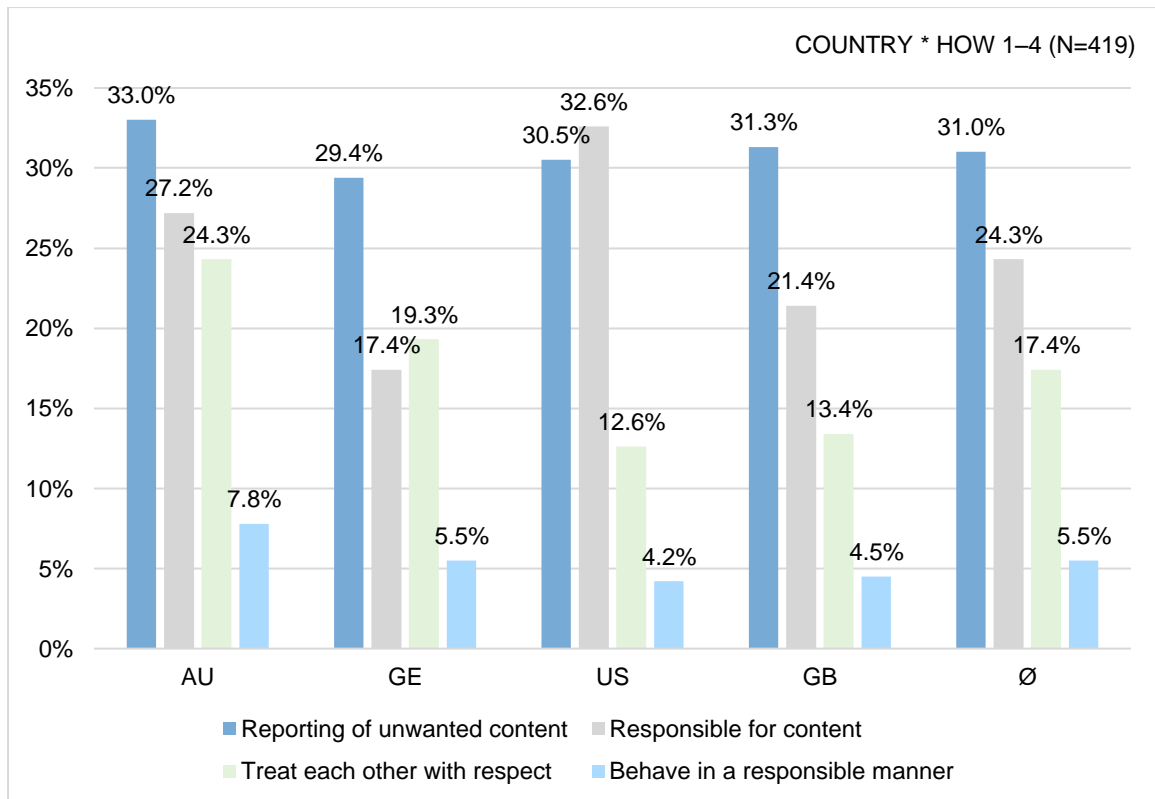


Abbildung 18: Länderspezifische Unterschiede beim von den Userinnen und Usern in den Richtlinien erwünschten Verhalten (in %)

Bei den unterschiedlichen Möglichkeiten, wie Userinnen und User regelwidrige Inhalte melden können, konnten einige signifikante Unterschiede zwischen den Ländern beobachtet werden. Der erste signifikante Unterschied konnte bei der Option, Inhalte zu markieren und dadurch direkt zu melden, festgestellt werden:  $X^2(6, N=415) = 12,583, p=0,050$ . Hier war die Stärke des Zusammenhangs mit 0,123 gering. In Deutschland haben Userinnen und User in 76,2% der Fälle die Option, Inhalte durch das Markieren direkt zu melden. In Österreich (68,0%), Großbritannien (59,8%) oder den Vereinigten Staaten (54,7%) wird diese Möglichkeit deutlich seltener angeboten. Ein weiterer signifikanter Unterschied konnte bei der Möglichkeit, Verstöße via Telefon zu melden, beobachtet werden:  $X^2(6, N=419) = 12,614, p=0,050$ . Hier war die Stärke des Zusammenhangs mit 0,123 ebenfalls gering. In Österreich und Deutschland wurde diese Möglichkeit in keiner einzigen Richtlinie angeboten. In Großbritannien können Userinnen und User immerhin in 5,4% der Fälle Verstöße an die Plattform-Betreiber via Telefon melden. Der letzte signifikante Unterschied konnte hinsichtlich der Meldung von Verstößen via Post festgestellt werden:  $X^2(6, N=419) = 13,419, p=0,037$ . Hier war die Stärke des Zusammenhangs mit 0,127 jedoch wieder gering. Durchschnittlich bieten 2,6% aller

Unternehmen ihren Userinnen und Usern die Möglichkeit, Verstöße via Post zu melden. In Großbritannien wird diese Option mit 7,2% deutlich häufiger angeboten als in Österreich (2,0%), Deutschland (0,9%) oder den Vereinigten Staaten, wo diese Möglichkeit gar nicht angeboten wird.

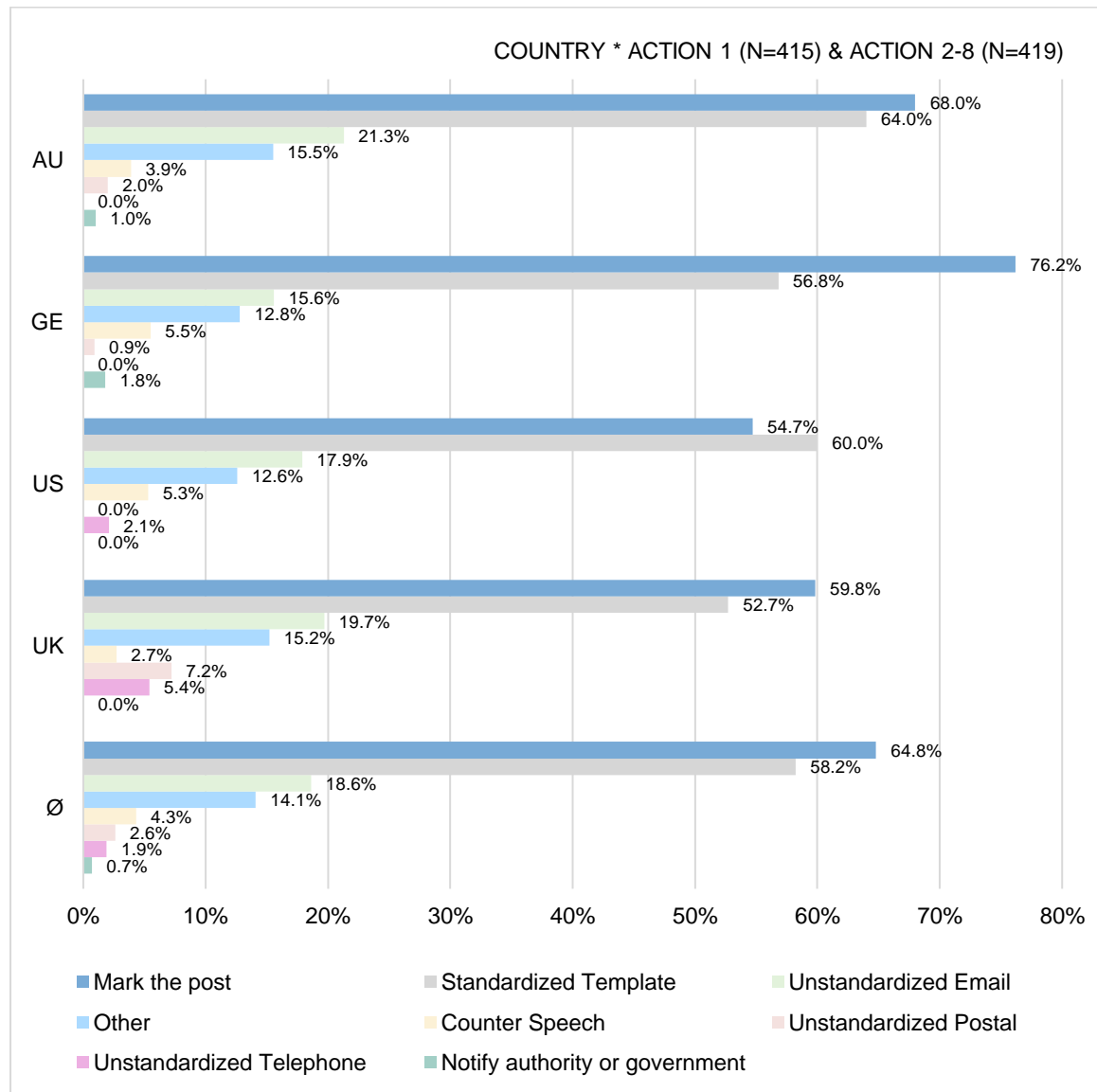


Abbildung 19: Länderspezifische Unterschiede in den Handlungsmöglichkeiten der Userinnen und User in Bezug auf HOC (in %)

Durchschnittlich enthalten 22,4% aller Richtlinien eine detaillierte Anleitung für die Meldung von regelwidrigen Inhalten. Aus Abbildung 20 geht hervor, dass Anleitungen zum Melden von HOC in Richtlinien aus Großbritannien mit 26,8% am häufigsten vorkommen.

Deutlich seltener, nämlich mit 19,3%, kommen derartige Anleitungen in deutschen Unternehmensrichtlinien vor. Hier ist allerdings kein signifikanter Unterschied zu beobachten.

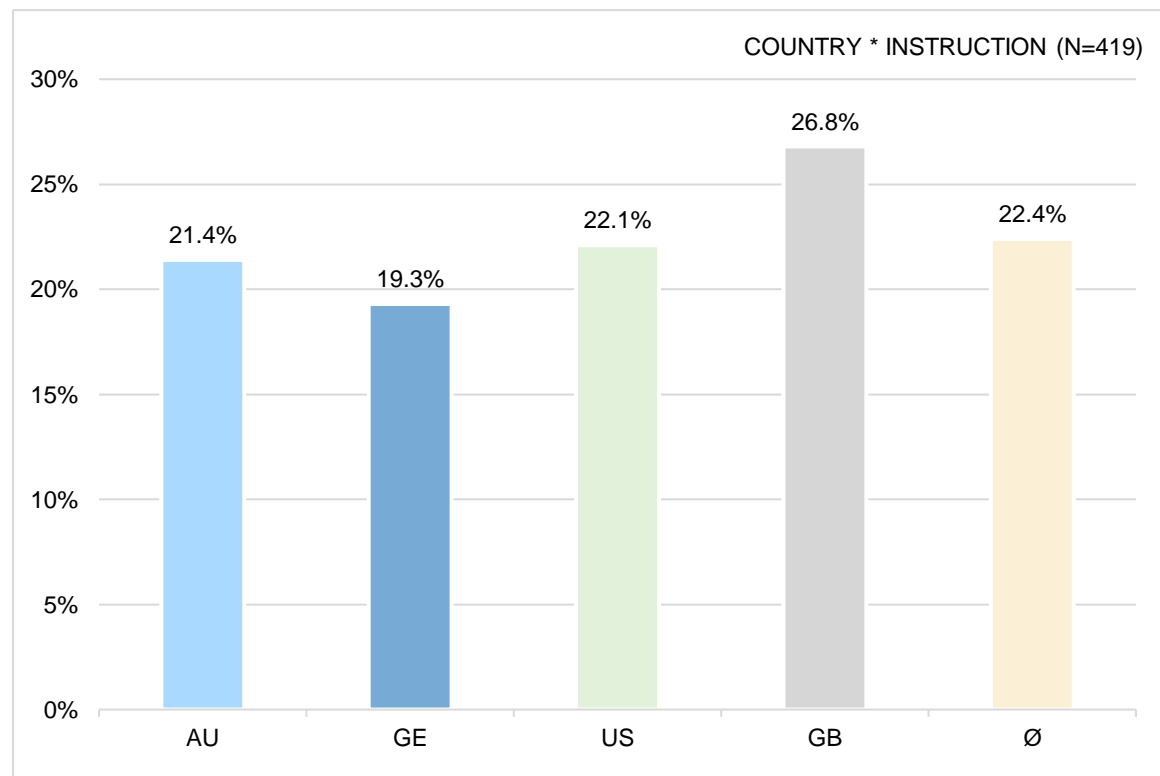


Abbildung 20: Anteil der Richtlinien, die eine detaillierte Anleitung zum Melden von HOC beinhalten, nach Land (in %)

Wie in Abbildung 21 ersichtlich ist, wird in durchschnittlich 18,1% der Richtlinien explizit erwähnt, welche Schritte von den Plattform-Betreibern eingeleitet werden, um einen gemeldeten Inhalt zu überprüfen. Aus den analysierten Richtlinien geht hervor, dass in Großbritannien die Userinnen und User mit 19,6% am häufigsten erfahren, was mit ihren gemeldeten Inhalten passiert. In den Vereinigten Staaten (18,9%), Deutschland (17,4%) und Österreich (16,5%) wird den Userinnen und Usern eher seltener mitgeteilt, was mit ihren gemeldeten Inhalten passiert. Auch hier gibt es vergleichsweise nur geringe länderspezifische Unterschiede, die allerdings nicht statistisch signifikant sind.

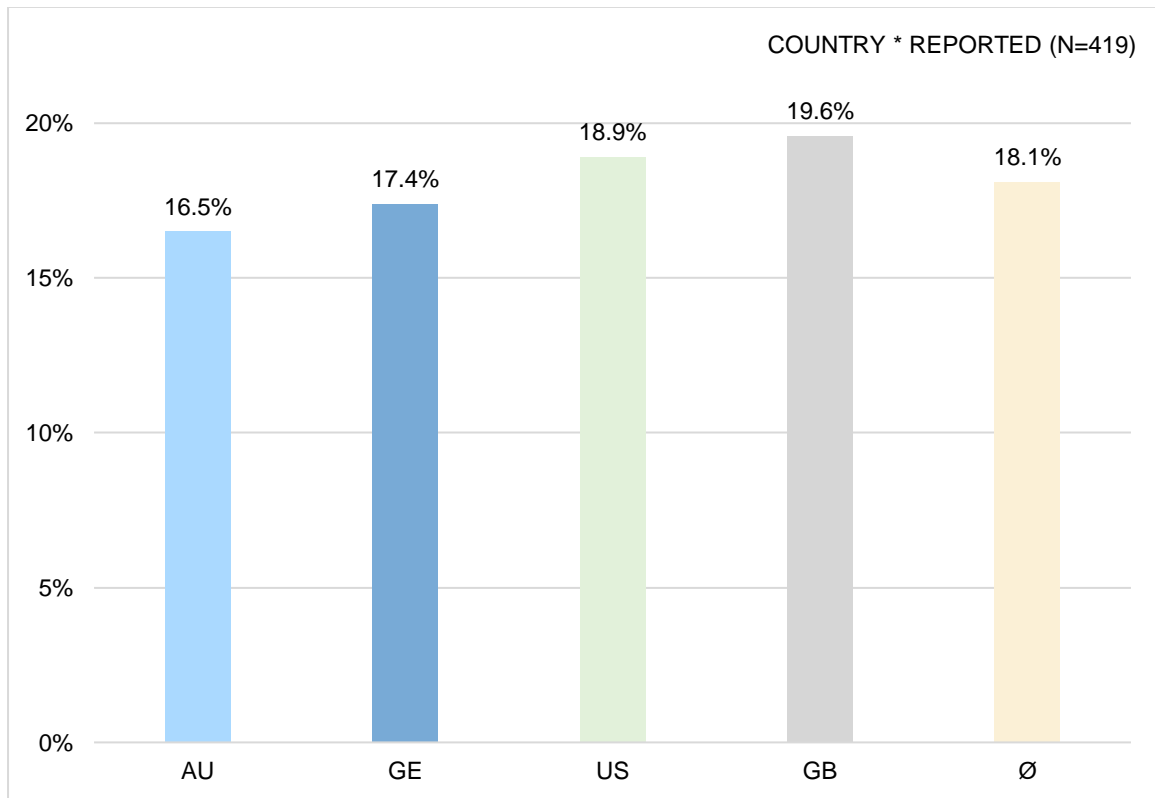


Abbildung 21: Anteil der Richtlinien, in denen erwähnt wird, wie die Unternehmen mit den gemeldeten Inhalten umgehen, nach Land (in %)

Hinsichtlich der unterschiedlichen Gründe für ein Verbot von HOC konnte ein signifikanter Unterschied festgestellt werden. In Abbildung 22 ist zu sehen, dass in Österreich mit 16,5% signifikant häufiger Verhaltensregeln für ein verantwortungsbewusstes bzw. ethisches Verhalten aufgestellt werden als in anderen Ländern:  $X^2(3, N=419) = 13,949$ ,  $p = 0,003$ . Die Stärke des Zusammenhangs ist hier allerdings mit einem Wert von 0,182 wie bereits zuvor eher gering. Plattform-Betreiber aus Österreich sehen sich mit 16,5% eher verpflichtet, trotz des hohen Stellenwerts der freien Meinungsäußerung gewisse Inhalte oder Verhaltensweisen zu verbieten, als Unternehmen aus den Vereinigten Staaten (11,6%) oder Großbritannien (9,8%). Die Förderung von konstruktiven Debatten wird in österreichischen Richtlinien mit 13,6% häufiger erwähnt als in Großbritannien (7,1%), den Vereinigten Staaten (6,3%) oder Deutschland (5,5%). Die länderspezifischen Unterschiede sind hier allerdings nicht signifikant.

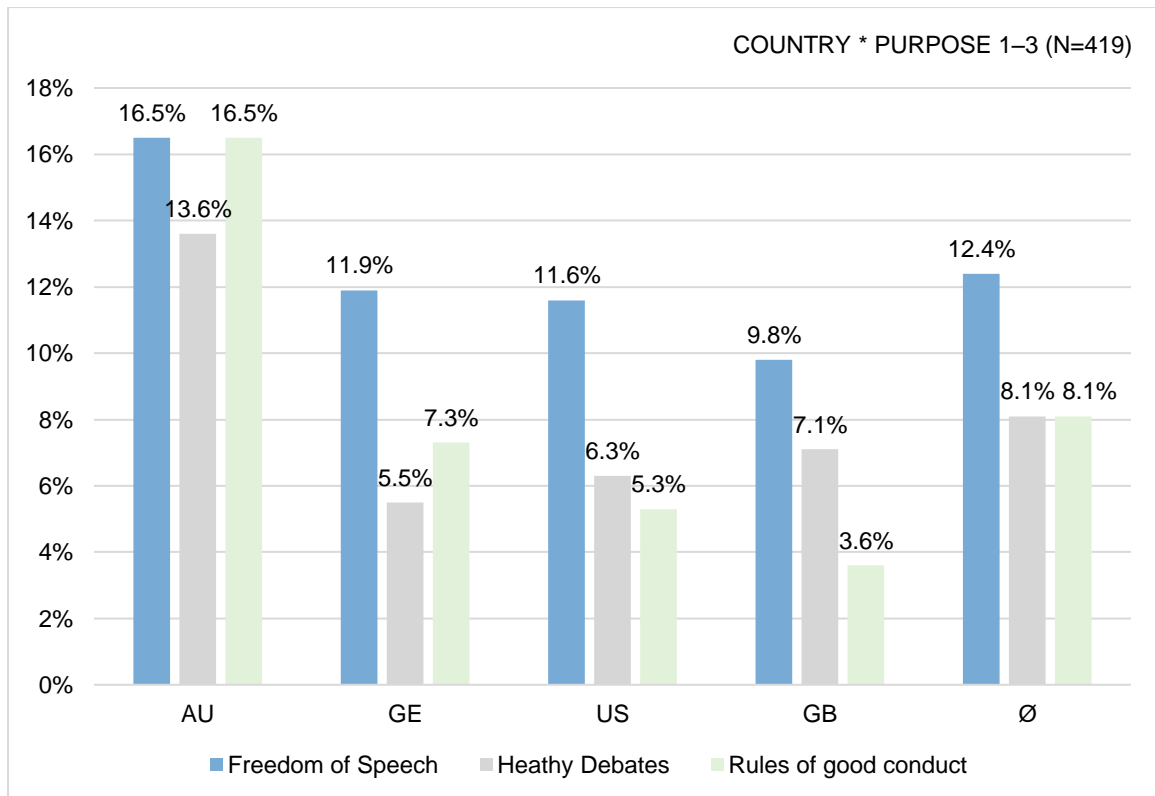


Abbildung 22: Länderspezifische Unterschiede hinsichtlich der Gründe für ein Verbot von HOC

Die länderspezifischen Unterschiede hinsichtlich der unterschiedlichen Konsequenzen bei einem Verstoß gegen die Richtlinien sind in Abbildung 23 dargestellt. Hier konnte ein signifikanter Unterschied beobachtet werden. In Österreich werden Userinnen und User, die regel- oder rechtswidrige Inhalte veröffentlichen, mit 35,0% wesentlich häufiger verklagt bzw. gerichtlich verfolgt als in anderen Ländern:  $X^2(3, N=419) = 9,367, p = 0,025$ . Die Stärke des Zusammenhangs ist hier mit einem Wert von 0,150 eher gering. In Großbritannien droht den Userinnen und Usern diese Konsequenz vergleichsweise nur in 17,0%. Bei den anderen Konsequenzen konnten nur geringe länderspezifische Unterschiede festgestellt werden, die nicht statistisch relevant waren. Die am häufigsten getroffene Maßnahme des Löschens von Inhalten ohne Angabe von Gründen wird in den Vereinigten Staaten am häufigsten getroffen, gefolgt von Österreich (67,0%), Deutschland (64,2%) und Großbritannien (62,5%). Das Löschen bzw. Sperren eines Accounts werden den Userinnen und Usern aus Österreich mit 61,2% am häufigsten angedroht. In Deutschland droht diese Konsequenz hingegen nur in 51,4% der untersuchten Fälle. Plattform-Betreiber aus Österreich warnen ihre Userinnen und User noch am ehesten (17,5%), bevor sie weitere Schritte einleiten. In Deutschland (17,4%), Großbritannien (13,4%) oder den Vereinigten Staaten (12,6%) werden die Userinnen und User etwas

seltener verwarnt. In Deutschland werden regelwidrige Inhalte bei der Löschung mit 14,7% am häufigsten mit einer Erklärung versehen. Diese Maßnahme wird in den anderen Ländern deutlich seltener getroffen (GB: 9,8%, US: 9,5%, AU: 8,7%). Eine weitere Konsequenz, die den Userinnen und Usern bei einem Verstoß gegen die Richtlinien droht, ist das Löschen bzw. Schließen der gesamten Diskussion. Diese Maßnahme wird allerdings eher selten angewandt. Am ehesten werden Diskussionen mit 7,1% noch in Großbritannien gelöscht bzw. geschlossen.

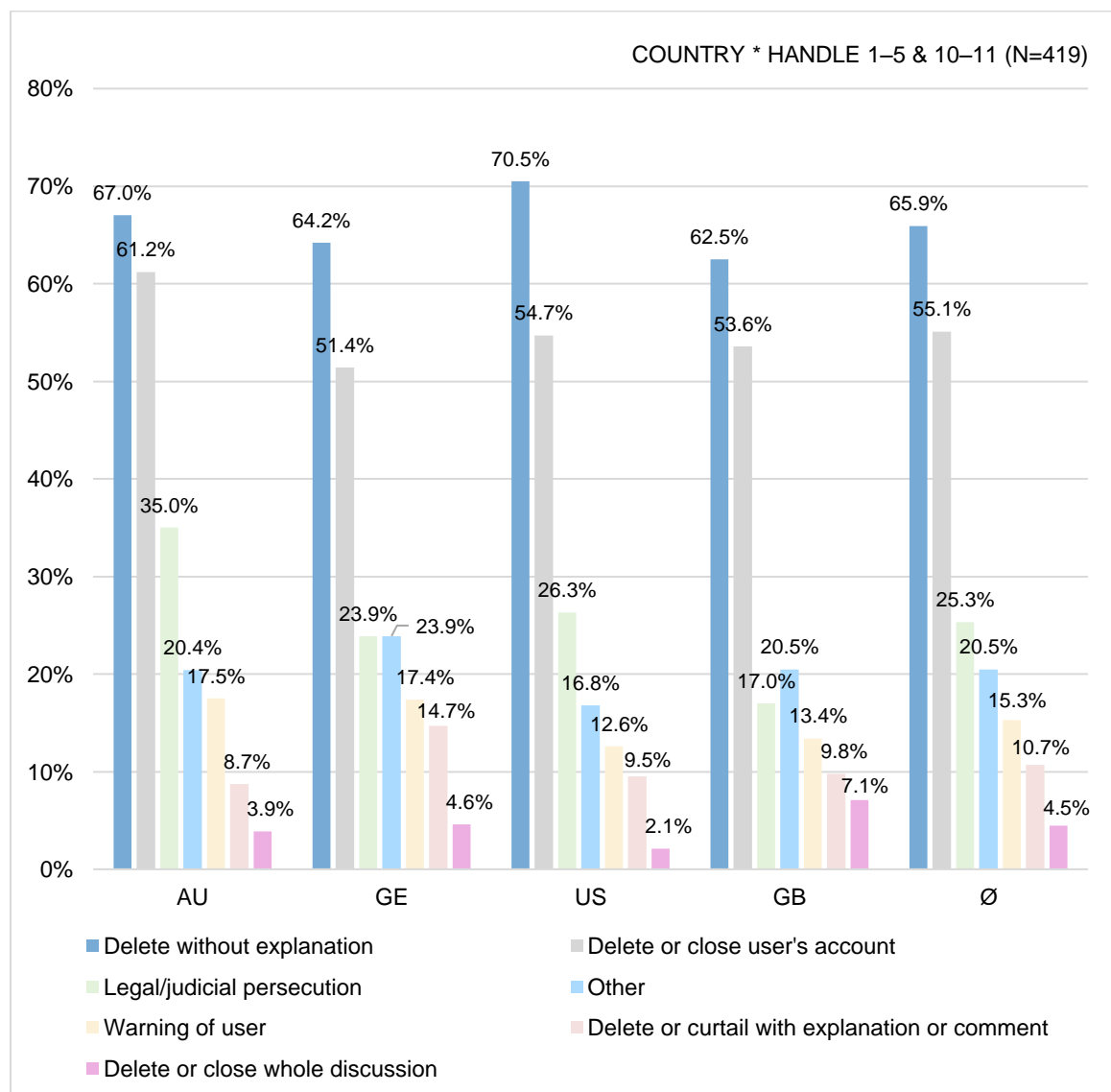


Abbildung 23: Länderspezifische Unterschiede hinsichtlich der Konsequenzen beim Verstoß gegen die Richtlinien

### 5.3. Diskussion der Ergebnisse

Durch die detaillierte Inhaltsanalyse von Unternehmensrichtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien ist es gelungen, Antworten auf die eingangs gestellten Forschungsfragen zu erhalten. Es konnten neue wertvolle Erkenntnisse gewonnen werden, die nun in diesem Kapitel besprochen werden.

Wie bereits im theoretischen Teil dieser Arbeit erwähnt wurde, ist es äußerst wichtig, bestimmte Maßnahmen zu treffen, um für ein faires, freundliches und gesundes Diskussionsklima zu sorgen. Aufgrund der Aktualität des Themas fühlen sich dementsprechend bereits viele Unternehmen verpflichtet, Verhaltensrichtlinien für ihre Online-Plattformen zu erstellen. Diese sollen unter anderem regeln, wie sich die Userinnen und User zu verhalten haben, welche Inhalte bzw. Verhaltensweisen verboten sind und welche Konsequenzen bei einem Verstoß gegen die Richtlinien drohen. Einerseits sollen Userinnen und User durch die Richtlinien aufgefordert werden, aktiv und engagiert gegen HOC vorzugehen, beispielsweise durch das Melden von regel- bzw. rechtswidrigen Inhalten oder in Form von Counter Speech. Andererseits sollen diese Richtlinien auch informieren, weshalb bestimmte Inhalte auf den Online-Plattformen nicht erwünscht sind. Vielen Personen ist nämlich nicht bewusst, welchen Schaden ihre Worte anrichten können. Deswegen ist es wichtig zu kommunizieren, dass beleidigende, aggressive und destruktive Äußerungen oder Inhalte nicht toleriert werden.

In Österreich wird beispielsweise besonders viel Wert darauf gelegt, beleidigende, einschüchternde, drohende oder illegale Äußerungen und Inhalte zu verbieten. Hierzu zählen auch Mobbing und Verleumdungen. Das könnte daran liegen, dass in Österreich üble Nachreden, Verhetzungen, Beleidigungen, Drohungen oder aber auch Cyber-Mobbing strafrechtlich unter Strafe stehen. Weiters werden in vielen Richtlinien auch pornographische und obszöne Inhalte verboten. Hierzu gibt es zwar kein Gesetz, dennoch wird diese Maßnahme zum Teil auch aus Kinder- und Jugendschutzgründen getroffen. Das Verbot der Diskriminierung wird im Vergleich zu anderen Inhalten eher selten erwähnt. Nur die Verbote von Hate Speech und Datenschutzverletzungen werden noch seltener erwähnt. Österreichische Unternehmen betonen in ihren Richtlinien besonders häufig, welche Verhaltensweisen sie von ihren Userinnen und Usern erwarten. Vor allem das Melden von regelwidrigen Inhalten, der respektvolle Umgang miteinander und das verantwortungsbewusste Verhalten sind in vielen österreichischen Richtlinien erwünscht. Um den Userinnen und Usern das Melden von regel- und rechtswidrigen Inhalten zu



erleichtern, werden ihnen hierfür unterschiedliche Möglichkeiten angeboten. Im Vergleich zu anderen Ländern erfahren in Österreich Userinnen und User eher selten, was mit ihren gemeldeten Inhalten passiert. Diesbezüglich gibt es noch Handlungsbedarf, da es für jene Personen, die regelwidrige Inhalte aufspüren und melden, durchaus interessant ist zu erfahren, wie mit ihren gemeldeten Inhalten umgegangen wird. Österreichische Plattform-Betreiber bemühen sich auch, ihren Userinnen und Usern zu erklären, weshalb bestimmte Inhalte nicht toleriert werden. Im Vergleich zu den Richtlinien aus Deutschland, den Vereinigten Staaten und Großbritannien thematisieren Richtlinien aus Österreich besonders häufig die Grenzen der freien Meinungsäußerung. Außerdem kommunizieren sie auch überdurchschnittlich oft, dass die verschriftlichten Richtlinien gesunde Debatten sowie ein ethisches und verantwortungsbewusstes Verhalten fördern sollen. Verstoßen Userinnen und User gegen die Unternehmensrichtlinien, drohen ihnen je nach Schwere des Verstoßes unterschiedliche Konsequenzen. Der *Kurier*, der durchaus als Good-Practice-Beispiel genannt werden kann, löscht primär Beiträge, die gegen die Richtlinien verstoßen. Darüber hinaus behält er sich aber auch das Recht vor, Diskussionsteilnehmerinnen und Diskussionsteilnehmer aus den Diskussionen auszuschließen und deren Nutzerdaten an Behörden weiterzuleiten, sofern sie offensichtlich rechtswidrige Inhalte verbreitet haben.

Aus den untersuchten Unternehmensrichtlinien aus Deutschland geht hervor, dass von allen Inhalten beleidigende und anstößige Äußerungen am häufigsten verboten werden. Illegale Äußerungen sowie Gewaltdarstellungen werden ebenfalls in vielen Richtlinien verboten. In Deutschland werden neben der Beleidigung und der üblen Nachrede nämlich auch unter anderem das Aufrufen zu öffentlichen Straftaten oder die Verbreitung von öffentlichen Gewaltdarstellungen unter Strafe gestellt. Das Verbot von pornographischen oder obszönen Inhalten, diskriminierenden Äußerungen, Datenschutzverletzungen oder Hate Speech wird zwar auch, aber weitaus seltener, erwähnt. Im Vergleich zu den untersuchten Richtlinien aus den anderen Ländern wird in Deutschland überdurchschnittlich häufig das Verbreiten von bestimmten Inhalten aus Kinder- und Jugendschutzgründen verboten. Deutsche Unternehmen ermutigen ihre Userinnen und User auch, für ein angenehmes Diskussionsklima zu sorgen. Hierfür bitten sie sie vor allem, regelwidrige Beiträge zu melden und respektvoll miteinander umzugehen. In 42,2% der Richtlinien werden auch Beispiele angeführt, wie die Userinnen und User auf den Online-Plattformen interagieren sollen. Dass vor allem größere Unternehmen besonders viel Wert darauf legen, ihre Community zum Melden von regel- bzw. rechtswidrigen Inhalten aufzufordern, liegt wohl daran, dass sie gesetzlich dazu verpflichtet sind,

offensichtlich rechtswidrige innerhalb von 24 Stunden zu entfernen. Andernfalls können sie für diese Inhalte haftbar gemacht werden. Um das Melden von derartigen Inhalten zu erleichtern, bieten auch die deutschen Plattform-Betreiber unterschiedliche (technische) Möglichkeiten an. Im Vergleich zu anderen Ländern werden in Deutschland allerdings nur selten detaillierte Anleitungen zum korrekten Melden von Beiträgen veröffentlicht. Interessant ist, dass Plattform-Betreiber in Deutschland ihre Userinnen und User besonders häufig dazu ermutigen, in Form von Counter Speech gegen HOC vorzugehen. Wie am Beispiel des deutschen Internetportals Web.de zu sehen ist, bemühen sich auch viele deutsche Unternehmen zu kommunizieren, weshalb gewisse Inhalte auf ihren Plattformen nicht toleriert werden können. Web.de erklärt, dass gewisse Regeln für eine rege, interessante und rechtskonforme Diskussion unabdingbar sind und dass gewisse Inhalte nicht durch das Recht auf freie Meinungsäußerung geschützt sind. Beiträge, die gegen die Richtlinien verstoßen, werden in den meisten Fällen ohne Angabe von Gründen gelöscht. Aus einigen Richtlinien, wie beispielsweise der Netiquette von *Zeit Online*, geht außerdem hervor, dass regelwidrige Beiträge mit einer Anmerkung versehen werden. Auch Kommentare, die zwar nicht gelöscht aber dennoch gekürzt werden, werden an der entsprechenden Stelle gekennzeichnet. Diese Maßnahme ist zwar zeitintensiv aber durchaus lobenswert, da sie für mehr Transparenz sorgt.

Die Vereinigten Staaten sind zwar große Verfechter der Meinungs- bzw. Redefreiheit, dennoch werden gewisse Äußerungen bzw. Inhalte unter Strafe gestellt. Darunter fallen unter anderem Obszönitäten, Verleumdungen und Beleidigungen. Aus der Studie des Pew Research Center (2015b) ging hervor, dass die Beschränkung von beleidigenden und anstößigen Aussagen gegenüber Minderheiten von Personen in europäischen Ländern eher toleriert werden als in den Vereinigten Staaten. Aus den für diese Forschungsarbeit analysierten Richtlinien geht hervor, dass diskriminierende Inhalte vergleichsweise zwar eher selten verboten werden, beleidigende und anstößige Äußerungen hingegen in den Richtlinien aus den Vereinigten Staaten am häufigsten verboten werden. Eine interessante Erkenntnis ist auch, dass viele Betreiber kommunizieren, dass die Userinnen und User ihrer Plattformen für ihre eigenen Inhalte verantwortlich sind, obwohl die Unternehmen in den Vereinigten Staaten nicht für fremde rechtswidrige Inhalte haften. Obwohl die Betreiber der Online-Plattformen gesetzlich nicht zur Löschung von regel- und rechtswidrigen Inhalten verpflichtet sind, wird diese Maßnahme am häufigsten in den Vereinigten Staaten getroffen. Verwarnungen von Userinnen und Usern, die gegen die Regeln verstoßen, oder das Löschen von Inhalten mit der Angabe von Gründen, sind hingegen eher nicht üblich. Ein Good-Practice-Beispiel aus den Vereinigten Staaten ist YouTube. Das Unternehmen bemüht sich um möglichst

viel Transparenz sowohl bei der visuellen und inhaltlichen Aufbereitung der Richtlinien als auch im Umgang mit den gemeldeten Inhalten. Personen, die regelwidrige Inhalte an YouTube melden wollen, erhalten hierfür eine detaillierte Anleitung. Außerdem werden sie darüber informiert, was mit den gemeldeten Inhalten passiert.

In Großbritannien gibt es derzeit noch keine Vorschriften, die Unternehmen verpflichten, auf eine bestimmte Art und Weise mit HOC umzugehen. Betreiber von Online-Plattformen orientieren sich hinsichtlich der verbotenen Inhalte an den rechtlichen Vorschriften, die unter anderem Diskriminierungen und beleidigende, drohende oder obszöne Äußerungen und Inhalte verbieten. Bestimmte Kinder- oder Jugendschutzmaßnahmen werden, im Vergleich zu anderen Ländern, in Großbritannien eher selten erwähnt. Auch der erwünschte respektvolle Umgang und ein verantwortungsbewusstes Verhalten auf den Online-Plattformen werden kaum thematisiert. Hier besteht durchaus Handlungsbedarf, da diese Maßnahmen maßgeblich dazu beitragen können, für ein besseres Diskussionsklima zu sorgen. Im Vergleich zu anderen Ländern haben die Userinnen und User in Großbritannien nicht nur die Möglichkeit, regelwidrige Beiträge zu markieren oder per Mail auf etwaige Verstöße gegen die Richtlinien hinzuweisen. In einigen Fällen können die Betreiber auch per Post oder Telefon kontaktiert werden. Vorbildlich sind die Unternehmen in Großbritannien nicht nur hinsichtlich der unterschiedlichen Möglichkeiten, wie Userinnen und User gegen HOC vorgehen können. In den meisten Richtlinien sind auch detaillierte Anleitungen vorzufinden, die das Melden von regelwidrigen Inhalten erleichtern sollen. Außerdem erwähnen die Unternehmen in Großbritannien überdurchschnittlich oft, wie sie mit den gemeldeten Inhalten umgehen. Zu den Konsequenzen, die bei einem Verstoß gegen die Richtlinien drohen, zählen am häufigsten das Löschen der Inhalte ohne Angabe von Gründen (obwohl keine Verpflichtung zur Löschung von regelwidrigen Inhalten besteht) und das temporäre oder dauerhafte Sperren der Userinnen und User. Als Good-Practice-Beispiel kann hier der Filehosting-Dienst Imgur genannt werden. In den Richtlinien wird explizit erklärt, welche Konsequenzen bei welcher Schwere des Verstoßes drohen. Um Verstöße möglichst gering zu halten, werden die verbotenen Inhalte mit Bildern versehen. So können die Userinnen und User besser erkennen, welche Inhalte geduldet werden und welche gegen die Richtlinien verstoßen.



## VI. Fazit

Die vorliegende Forschungsarbeit ist der Frage nachgegangen, wie Unternehmen in Österreich, Deutschland, den Vereinigten Staaten und Großbritannien mit Harmful Online Communication umgehen. Anhand einer Inhaltsanalyse von insgesamt 419 Richtlinien wurde erforscht, welche Inhalte laut der Unternehmensrichtlinien zu HOC zählen, welche Maßnahmen von den Unternehmen getroffen werden, um gegen HOC vorzugehen und ob es länderspezifische Unterschiede in Bezug auf HOC gibt.

Zusammenfassend lässt sich festhalten, dass die Unternehmen ihrer sozialen Verpflichtung, HOC auf ihren Plattformen einzudämmen, zunehmend nachkommen. In ihren Richtlinien kommunizieren sie, dass sie gewisse Inhalte wie unter anderem Beleidigungen, Verspottungen, Diskriminierungen, Hate Speech oder Pornographie auf ihren Plattformen nicht tolerieren. Aus den Richtlinien geht auch hervor, dass die Unternehmen bevorzugt restriktive Maßnahmen anwenden, um gegen HOC vorzugehen. Inhalte, die gegen die Richtlinien verstoßen, werden in den meisten Fällen umgehend gelöscht. Viele Unternehmen wollen aber auch HOC mit Hilfe ihrer Community eindämmen. In vielen Richtlinien werden Userinnen und User deshalb gebeten, regel- und rechtswidrige Inhalte zu melden. Userinnen und User können aber auch in Form von Counter Speech selbstständig gegen HOC vorgehen. Hierzu werden sie in den untersuchten Richtlinien allerdings nur selten aufgefordert. Diesbezüglich besteht durchaus noch Handlungsbedarf, da HOC mit restriktiven Maßnahmen nur oberflächlich beseitigt werden kann. Plattform-Betreiber sollten ihre Userinnen und User vermehrt zum aktiven Handeln motivieren. Betroffene aber auch unbeteiligte Dritte sollten HOC nicht einfach ignorieren, sondern sich aktiv dagegen aussprechen, um HOC nicht salonfähig zu machen. Eine aufgeklärte, sensibilisierte Community kann nämlich durch aktives Handeln maßgeblich zu einem positiveren und konstruktiveren Diskussionsklima beitragen.

In Bezug auf die länderspezifischen Unterschiede lässt sich festhalten, dass die untersuchten Richtlinien aus Österreich, Deutschland, den Vereinigten Staaten oder Großbritannien keine großen Unterschiede aufweisen. Die Richtlinien sind sich inhaltlich überwiegend ähnlich. Es konnten nur einige wenige signifikante Unterschiede festgestellt werden. So wurde beispielsweise in Richtlinien aus Deutschland häufiger als in Richtlinien aus Großbritannien erwähnt, wie die Userinnen und User miteinander interagieren sollen. Weitere Unterschiede konnten hinsichtlich der unterschiedlichen Handlungsmöglichkeiten der Userinnen und User in Bezug auf HOC beobachtet werden. In Deutschland konnten

nämlich regelwidrige Inhalte weitaus häufiger markiert und direkt gemeldet werden als in den Vereinigten Staaten. Außerdem wurde die Option, Verstöße via Telefon oder Post an die Plattform-Betreiber zu melden, in Großbritannien häufiger angeboten als in den anderen Ländern. Schlussendlich konnte auch festgestellt werden, dass in Österreich signifikant häufiger Verhaltensregeln für ein verantwortungsbewusstes bzw. ethisches Verhalten („rules of good conduct“) aufgestellt wurden als in den anderen Ländern.

In dieser Forschungsarbeit wurde nicht erhoben, ob sich die Userinnen und User registrieren müssen, um auf den jeweiligen Online-Plattformen kommentieren zu können. Es wäre allerdings durchaus interessant herauszufinden, ob Unternehmen nun vermehrt dem Beispiel der *Huffington Post* folgen und anonyme Postings ebenfalls verbieten, um gegen aggressive, destruktive oder hasserfüllte Äußerungen vorzugehen. Diese Maßnahme kann nämlich durch die erleichterte Identifizierbarkeit nachweislich zu einem höflicheren Umgangston verhelfen. Zukünftige Forschungsarbeiten könnten auch unterschiedliche Maßnahmen, wie die Prä- und Postmoderation oder die verpflichtete Registrierung miteinander vergleichen, um herauszufinden, wie HOC am effektivsten eingedämmt werden kann. Auf diesem Gebiet besteht jedenfalls noch ausreichend Handlungs- und Forschungsbedarf, da davon ausgegangen werden kann, dass Hass im Netz auch in Zukunft ein weitreichendes gesellschaftliches Problem bleiben wird.

## VII. Quellenverzeichnis

- Amerika Dienst. (2016). *Meinungsfreiheit in den Vereinigten Staaten*. Abgerufen am 14. Juli 2019 von <https://de.usembassy.gov/de/meinungsfreiheit-4/>
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "Nasty Effect": Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication* (19), S. 373-387.
- Baldauf, J., Banaszczuk, Y., Koreng, A., Schramm, J., & Stefanowitsch, A. (2015). *Die direkte Bedrohung durch Hate Speech darf nicht unterschätzt werden! Interview mit Dorothee Scholz, Diplompsychologin*. Abgerufen am 5. Juli 2019 von „Geh sterben!“ Umgang mit Hate Speech und Kommentaren im Internet: <https://www.amadeu-antonio-stiftung.de/wp-content/uploads/2018/08/hatespeech-1.pdf>
- Banks, J. (2010). Regulation hate speech online. *International Review of Law, Computer & Technology*, 24(3), S. 233-239.
- Bezemek, C. (2015). *Freie Meinungsäußerung: Strukturfragen des Schutzgegenstandes im Rechtsvergleich zwischen dem Ersten Zusatz zur US Verfassung und Artikel 10 der Europäischen Menschenrechtskonvention*. Verlag Österreich.
- BGBl. I. (2007). *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken vom 1. September 2017*. Abgerufen am 13. Juli 2019 von [https://www.bgbl.de/xaver/bgbl/start.xav#\\_\\_bgbl\\_\\_%2F%2F\\*%5B%40attr\\_id%3D%27bgbl117s3352.pdf%27%5D\\_\\_1563017670182](https://www.bgbl.de/xaver/bgbl/start.xav#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl117s3352.pdf%27%5D__1563017670182)
- Blunt, James (@JamesBlunt). (30. 04. 2014). Abgerufen am 19. Juli 2019 von Then you need to see a doctor Tweet: <https://twitter.com/jamesblunt/status/461640897030275073?lang=de>
- Brodnig, I. (2016). *Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können*. Brandstätter.
- Brosius, H.-B., Koschel, F., & Haas, A. (2008). *Methoden der empirischen Kommunikationsforschung. Eine Einführung. 4., überarbeitete und erweiterte Auflage*. VS Verlag für Sozialwissenschaften.
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), S. 297-326.
- Buckels, J., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. In *Personality and Individual Differences* (S. 97-102). Amsterdam: Elsevier.

- Cho, D., & Acquisti, A. (2013). *The More Social Cues, The Less Trolling? An Empirical Study of Commenting Behavior*. Abgerufen am 19. Juli 2019 von <https://www.econinfosec.org/archive/weis2013/papers/ChoWEIS2013.pdf>
- Civil Comments. (2016). *Publisher Documentation*. Abgerufen am 18. Juli 2019 von <https://civil.gitbooks.io/publisher-docs/content/>
- Code of Conduct on Countering Illegal Hate Speech Online*. (2016). Abgerufen am 26. Juni 2019 von [http://ec.europa.eu/justice/fundamental-rights/files/hate\\_speech\\_code\\_of\\_conduct\\_en.pdf](http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf)
- Cohen-Almagor, R. (2017). Balancing Freedom of Expression and Social Responsibility on the Internet. *Philosophia*(45), S. 973-985.
- Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). *Automated Hate Speech Detection and the Problem of Offensive Language*.
- dejure.org. (o.J.a). *Strafgesetzbuch, § 111 Öffentliche Aufforderung zu Straftaten*. Abgerufen am 12. Juli 2019 von <https://dejure.org/gesetze/StGB/111.html>
- dejure.org. (o.J.b). *Strafgesetzbuch, § 130 Volksverhetzung*. Abgerufen am 12. Juli 2019 von <https://dejure.org/gesetze/StGB/130.html>
- dejure.org. (o.J.c). *Strafgesetzbuch, § 131 Gewaltdarstellung*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/131.html>
- dejure.org. (o.J.d). *Strafgesetzbuch, § 185 Beleidigung*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/185.html>
- dejure.org. (o.J.e). *Strafgesetzbuch, § 186 Üble Nachrede*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/186.html>
- dejure.org. (o.J.f). *Strafgesetzbuch, § 187 Verleumdung*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/187.html>
- dejure.org. (o.J.g). *Strafgesetzbuch, § 240 Nötigung*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/240.html>
- dejure.org. (o.J.h). *Strafgesetzbuch, § 241 Bedrohung*. Abgerufen am 13. Juli 2019 von <https://dejure.org/gesetze/StGB/241.html>
- derStandard.at. (2016a). *iPhones für "Asylanten": Caritas und Hartlauer dementieren Facebook-Posting*. Abgerufen am 3. Juli 2019 von <https://www.derstandard.at/story/2000042007393/iphone-fuer-asylanten-erneut>
- derStandard.at. (2016b). *Facebook hat eine Art journalistische Verantwortung*. Abgerufen am 16. Juli 2019 von <https://www.derstandard.at/story/2000048415535/facebook-hat-eine-art-journalistische-verantwortung>



- derStandard.at. (2016c). *Wie wir uns konstruktive Onlinedebatten vorstellen*. Abgerufen am 17. Juli 2019 von <https://www.derstandard.at/story/2000047750036/ansaetze-fuer-konstruktive-onlinedebatten>
- derStandard.at. (2018). *Forenregeln. Community-Richtlinien*. Abgerufen am 17. Juli 2019 von <https://www.derstandard.at/communityrichtlinien>
- derStandard.at. (2019a). *Facebook unterstützt Frankreichs Maßnahmen gegen Hassbotschaften*. Abgerufen am 1. Juli 2019 von <https://www.derstandard.at/story/2000105428804/facebook-unterstuetzt-frankreichs-massnahmen-gegen-hassbotschaften>
- derStandard.at. (2019b). *Was macht das Community-Management des STANDARD?* Abgerufen am 17. Juli 2019 von <https://www.derstandard.at/story/2000101149798/was-macht-das-community-management-des-standard>
- Deutscher Bundestag. (2017). *Drucksache 18/12356*. Abgerufen am 13. Juli 2019 von Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken: <http://dipbt.bundestag.de/dip21/btd/18/123/1812356.pdf>
- Deutscher Bundestag. (o.J.). *Die Grundrechte*. Abgerufen am 12. Juli 2019 von [https://www.bundestag.de/parlament/aufgaben/rechtsgrundlagen/grundgesetz/gg\\_01-245122](https://www.bundestag.de/parlament/aufgaben/rechtsgrundlagen/grundgesetz/gg_01-245122)
- Diakopoulos, N., & Naaman, M. (2011). *Towards Quality Discourse in Online News Comments. CSCW 2011*. Hangzhou, China.
- DiePresse.com. (2019). *Hasspostings gegen Ulli Sima: Erste Verurteilung*. Abgerufen am 11. Juli 2019 von [https://diepresse.com/home/panorama/wien/5628606/Hasspostings-gegen-Ulli-Sima\\_Erste-Verurteilung](https://diepresse.com/home/panorama/wien/5628606/Hasspostings-gegen-Ulli-Sima_Erste-Verurteilung)
- EGMR. (7. 12. 1976). *Handyside v. UK, 5493/72*. Abgerufen am 11. Juli 2019 von [https://hudoc.echr.coe.int/eng#{"itemid":\["001-57499"\]}](https://hudoc.echr.coe.int/eng#{)
- EGMR. (1. 07. 1997). *Oberschlick v. Austria, 20834/92*. Abgerufen am 11. Juli 2019 von [https://hudoc.echr.coe.int/eng#{"itemid":\["001-58044"\]}](https://hudoc.echr.coe.int/eng#{)
- Einwiller, S., & Kim, S. (2018). *Organizational efforts to prevent harmful online communication - A cross national analysis of online platform providers' policies*.
- Europäische Kommission. (31. 05 2016). *Europäische Kommission und IT-Unternehmen geben Verhaltenskodex zur Bekämpfung illegaler Hassrede im Internet bekannt*. Abgerufen am 5. Juni 2019 von [http://europa.eu/rapid/press-release\\_IP-16-1937\\_de.htm](http://europa.eu/rapid/press-release_IP-16-1937_de.htm)

- Europäische Kommission gegen Rassismus und Intoleranz. (2016). *Allgemeine Politik-Empfehlung Nr. 15 der ECRI über die Bekämpfung von Hassrede*. Abgerufen am 11. Juni 2019
- Europarat. (2019). *Charts of signatures and ratifications of Treaty 005. Convention for the Protection of Human Rights and Fundamental Freedoms*. Abgerufen am 7. Juli 2019 von [https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005/signatures?p\\_auth=pHWIcipN](https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005/signatures?p_auth=pHWIcipN)
- Europarat/Ministerkomitee. (1997). *Anhang zu Empfehlung Nr. R (97) 20 des Ministerkomitees an die Mitgliedstaaten über „Hassrede“*. Abgerufen am 11. Juni 2019 von <http://www.egmr.org/minkom/ch/rec1997-20.pdf>
- European Convention on Human Rights. (1950). *Convention for the Protection of Human Rights and Fundamental Freedoms*. Abgerufen am 7. Juli 2019 von [https://www.echr.coe.int/Documents/Convention\\_ENG.pdf](https://www.echr.coe.int/Documents/Convention_ENG.pdf)
- Facebook. (2019). *Gemeinschaftsstandards*. Abgerufen am 16. Juli 2019 von <https://www.facebook.com/communitystandards/introduction>
- Facebook Newsroom. (2018). *Hard Questions: Who Reviews Objectionable Content on Facebook — And Is the Company Doing Enough to Support Them?* Abgerufen am 19. Juli 2019 von <https://newsroom.fb.com/news/2018/07/hard-questions-content-reviewers/>
- Faz.net. (2016). *Meinungsfreiheit als Deckmantel*. Abgerufen am 12. Juli 2019 von <https://www.faz.net/aktuell/politik/urteil-gegen-facebook-hetzer-meinungsfreiheit-als-deckmantel-14162226.html#/elections>
- Faz.net. (2017). *Facebook-Nutzer muss 2000 Euro wegen Beleidigung zahlen*. Abgerufen am 13. Juli 2019 von <https://www.faz.net/aktuell/politik/inland/facebook-nutzer-zahlt-strafe-wegen-beleidigung-von-claudia-roth-14877596.html>
- Financial Times. (2019). *Facebook and Google join ad agencies and brands in ‘safety’ alliance*. Abgerufen am 15. Juli 2019 von <https://www.ft.com/content/cf7a21a0-9131-11e9-aea1-2b1d33ac3271>
- Früh, W. (2017). *Inhaltsanalyse. Theorie und Praxis*. UTB.
- Gelber, K. (2019). Differentiating hate speech: a systemic discrimination approach. *Critical Review of International Social and Political Philosophy*.
- Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, 22(3), S. 324-341.
- Grimm, P. (2016). *Digitale Wertekultur statt einer Kultur der Verachtung*. Abgerufen am 13. Juni 2018 von Grünbuch Digitale Courage. Im Auftrag des Präsidenten des Bundesrates Mario Lindner:

- [https://parlament.gv.at/ZUSD/PDF/Gruenbuch\\_Digitale\\_Courage\\_Republik\\_Oest\\_erreich\\_Bundesrat.pdf](https://parlament.gv.at/ZUSD/PDF/Gruenbuch_Digitale_Courage_Republik_Oest_erreich_Bundesrat.pdf)
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*(6), S. 215-242.
- Hardaker, C. (2013). "Uh....not to be nitpicky,,,,,but...the past tense of drag is dragged, not drug." An overview of trolling strategies. *Journal of Language Aggression and Conflict*, 1(1), S. 58-86.
- Ksiazek, T. B. (2015). Civil Interactivity: How News Organizations' Commenting Policies Explain Civility and Hostility in User Comments. *Journal of Broadcasting & Electronic Media* , S. 556-573.
- Landers, E. (22. 08. 2013). *Huffington Post to ban anonymous comments*. Abgerufen am 19. Juli 2019 von CNN Business:  
<https://edition.cnn.com/2013/08/22/tech/web/huffington-post-anonymous-comments/index.html>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. In *Computers in Human Behavior* (S. 434-443). Amsterdam: Elsevier.
- Legal Information Institute. (o.J.). *47 U.S. Code § 230. Protection for private blocking and screening of offensive material*. Abgerufen am 14. Juli 2019 von  
<https://www.law.cornell.edu/uscode/text/47/230>
- legislation.gov.uk. (2001). *Public Order Act 1986, Part III, Section 27*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/1986/64/section/27#1365595>
- legislation.gov.uk. (2004). *Human Rights Act 1998, Art. 10*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/1998/42/schedule/1/part/I/chapter/9>
- legislation.gov.uk. (2007). *Racial and Religious Hatred Act 2006, Section 1*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/2006/1/section/1>
- legislation.gov.uk. (2010). *Criminal Justice and Immigration Act 2008, Section 74*. Abgerufen am 15. Juli 2019 von  
<https://www.legislation.gov.uk/ukpga/2008/4/section/74>
- legislation.gov.uk. (2015a). *Malicious Communications Act 1988, Section 1*. Abgerufen am 15. Juli 2019 von  
<http://www.legislation.gov.uk/ukpga/1988/27/section/1#commentary-c803482>
- legislation.gov.uk. (2015b). *Communications Act 2003, Section 127*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/2003/21/section/127>

- legislation.gov.uk. (2017a). *Public Order Act 1986, Part III, Section 18*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/1986/64/part/III/2017-04-06>
- legislation.gov.uk. (2017b). *Public Order Act, Part I, Section 4 & 4A*. Abgerufen am 15. Juli 2019 von <http://www.legislation.gov.uk/ukpga/1986/64/part/I>
- LexisNexis. (1942). *Chaplinsky v. New Hampshire, 315 U.S. 568, 62 S. Ct. 766 (1942)*. Abgerufen am 14. Juli 2019 von <https://www.lexisnexis.com/lawschool/resources/p/casebrief-chaplinsky-v-new-hampshire.aspx>
- Müller, K., & Schwarz, C. (2018). *Fanning the Flames of Hate: Social Media and Hate Crime*. Abgerufen am 5. Juli 2019 von [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3082972](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3082972)
- Meibauer, J. (2013). Hassrede – von der Sprache zur Politik. In J. Meibauer (Hrsg.), *Hassrede / Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion* (S. 1-16). Universität Gießen Bibliothek.
- National Constitution Center. (o.J.). *Amendment I: Freedom of Religion, Speech, Press, Assembly, and Petition*. Abgerufen am 14. Juli 2019 von <https://constitutioncenter.org/interactive-constitution/amendments/amendment-i>
- Nationales Komitee No Hate Speech Österreich. (2018). *Empfehlungen des Nationalen No Hate Speech Komitees an die Bundesregierung und Landesregierungen*. Abgerufen am 5. Juni 2019 von [https://www.nohatespeech.at/wp-content/uploads/2018/08/Empfehlungen\\_No\\_Hate\\_Speech\\_Komitee.pdf](https://www.nohatespeech.at/wp-content/uploads/2018/08/Empfehlungen_No_Hate_Speech_Komitee.pdf)
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks: Sage Publications.
- OGH. (5. 04. 2017). *15 Os 128/16m*. Abgerufen am 11. Juli 2019 von [https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Justiz&Dokumentnummer=JJT\\_20170405\\_OGH0002\\_0150OS00128\\_16M0000\\_000](https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Justiz&Dokumentnummer=JJT_20170405_OGH0002_0150OS00128_16M0000_000)
- Pariser, E. (2011). *The Filter Bubble. What the Internet Is Hiding from You*. Penguin Books.
- parlament.gv.at. (2019). *Bundesgesetz über Sorgfalt und Verantwortung im Netz; KommAustria-Gesetz, Änderung (134/ME)*. Abgerufen am 17. Juli 2019 von [https://www.parlament.gv.at/PAKT/VHG/XXVI/ME/ME\\_00134/index.shtml](https://www.parlament.gv.at/PAKT/VHG/XXVI/ME/ME_00134/index.shtml)
- Pew Research Center. (2015a). *Global Support for Principle of Free Expression, but Opposition to Some Forms of Speech*. Abgerufen am 19. Juli 2019 von <https://www.pewresearch.org/global/2015/11/18/appendix-a-3/>
- Pew Research Center. (2015b). *40% of Millennials OK with limiting speech offensive to minorities*. Abgerufen am 19. Juli 2019 von <https://www.pewresearch.org/fact-tank/2015/11/20/40-of-millennials-ok-with-limiting-speech-offensive-to-minorities/>

- Rechtsinformationssystem Bundeskanzleramt Österreich. (2015a). *Bundesrecht konsolidiert: Strafgesetzbuch § 107, tagesaktuelle Fassung*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetze=10002296&Artikel=&Paragraf=107&Anlage=&Uebergangsrecht=>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2015b). *Bundesrecht konsolidiert: Strafgesetzbuch § 5*. Abgerufen am 12. Juli 2019 von <https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bundesnormen&Dokumentnummer=NOR12029546>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2015c). *Bundesrecht konsolidiert: § 16 E-Commerce-Gesetz*. Abgerufen am 12. Juli 2019 von <https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bundesnormen&Dokumentnummer=NOR40025812>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2015d). *Bundesrecht konsolidiert: Allgemeines bürgerliches Gesetzbuch § 1330*. Abgerufen am 19. Juli 2019 von <https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bundesnormen&Dokumentnummer=NOR12019074>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2018a). *Bundesrecht konsolidiert: Strafgesetzbuch § 111, tagesaktuelle Fassung*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetze=10002296&Artikel=&Paragraf=111&Anlage=&Uebergangsrecht=>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2018b). *Bundesrecht konsolidiert: Strafgesetzbuch § 115, tagesaktuelle Fassung*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetze=10002296&Artikel=&Paragraf=115&Anlage=&Uebergangsrecht=>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2018c). *Bundesrecht konsolidiert: Strafgesetzbuch § 283, tagesaktuelle Fassung*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetze=10002296&Artikel=&Paragraf=283&Anlage=&Uebergangsrecht=>
- Rechtsinformationssystem Bundeskanzleramt Österreich. (2018d). *Bundesrecht konsolidiert: § 6 MedienG*. Abgerufen am 12. Juli 2019 von

<https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bundesnormen&Dokumentnummer=NOR40064952>

Rechtsinformationssystem Bundeskanzleramt Österreich. (2019a). *Bundesrecht konsolidiert: Gesamte Rechtsvorschrift für E-Commerce-Gesetz, Fassung vom 02.07.2019*. Abgerufen am 1. Juli 2019 von <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20001703>

Rechtsinformationssystem Bundeskanzleramt Österreich. (2019b). *Bundesrecht konsolidiert: Gesamte Rechtsvorschrift für Staatsgrundgesetz über die allgemeinen Rechte der Staatsbürger, Fassung vom 11.07.2019*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10000006>

Rechtsinformationssystem Bundeskanzleramt Österreich. (2019c). *Bundesrecht konsolidiert: Strafgesetzbuch § 107c, Fassung vom 11.07.2019*. Abgerufen am 11. Juli 2019 von <https://www.ris.bka.gv.at/Dokument.wxe?Abfrage=Bundesnormen&Dokumentnummer=NOR40177258>

Rechtsinformationssystem Bundeskanzleramt Österreich. (2019d). *Bundesrecht konsolidiert: Gesamte Rechtsvorschrift für Verbotsgesetz 1947, Fassung vom 12.07.2019*. Abgerufen am 12. Juli 2019 von <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10000207>

Ruane, K. A. (2014). *Freedom of Speech and Press: Exceptions to the First Amendment*. Abgerufen am 14. Juli 2019 von <https://fas.org/sgp/crs/misc/95-815.pdf>

Saferinternet.at. (o.J.). *Gegenrede – Wie kann ich gegen Hasspostings argumentieren?* Abgerufen am 19. Juli 2019 von <https://www.saferinternet.at/faq/problematische-inhalte/gegenrede-wie-kann-ich-gegen-hasspostings-argumentieren/>

Schmitt, J. B. (2017). Online-Hate Speech: Definition und Verbreitungsmotivationen aus psychologischer Perspektive. In K. Kaspar, L. Gräßler, & A. Riffi, *Online Hate Speech: Perspektiven auf eine neue Form des Hasses* (S. 52-56).

Schnellenbach, J. (2018). *ORDO*, 68(1), S. 159–178.

Schwertner, K. (03. 01. 2018). *Mach auch du mit! Facebook-Post*. Abgerufen am 19. Juli 2019 von <https://www.facebook.com/klaus.schwertner/posts/10155602315809807>

- Sirsch, J. (2013). Die Regulierung von Hassrede in liberalen Demokratien. In J. Meibauer (Hrsg.), *Hassrede/Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion* (S. 165-193). Gießen: Gießener Elektronische Bibliothek.
- Soll, I. (2001). *Die Meinungsäußerungsfreiheit in den Streitkräften: Ein Rechtsvergleich zwischen der Bundesrepublik Deutschland und der Republik Österreich*. Münster: LIT.
- Sponholz, L. (2018). *Hate Speech in den Massenmedien. Theoretische Grundlagen und empirische Umsetzung*. Wiesbaden: Springer VS.
- Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, 7 (3), S. 321-326.
- Sumner, W. G. (2007). *Folkways. A Study of Mores, Manners, Customs and Morals*. Cosimo Classics.
- SZ.de. (2018). *Sind Facebook und die AfD schuld, wenn Flüchtlingsheime brennen?* Abgerufen am 6. Juli 2019 von <https://www.sueddeutsche.de/politik/soziale-medien-sind-facebook-und-die-afd-schuld-wenn-fluechtlingsheime-brennen-1.4101610>
- Tagesspiegel.de. (2018). *Die Hauptursache liegt in einem Gefühl der Bedrohung*. Abgerufen am 30. Juni 2019 von <https://www.tagesspiegel.de/politik/verrohtedebattenkultur-die-hauptursache-liegt-in-einem-gefuehl-der-bedrohung/22930552.html>
- The Telegraph. (2018). *British MPs call for German-style law to block hate speech on social media*. Abgerufen am 15. Juli 2019 von <https://www.telegraph.co.uk/technology/2018/07/28/british-mps-call-german-style-law-block-hate-speech-social-media/>
- Tsisis, A. (2009). *Dignity and Speech: The Regulation of Hate Speech in a Democracy*. 42 Wake Forest Law Review 497.
- Twitter. (2019). *The Twitter Rules*. Abgerufen am 17. Juli 2019 von <https://help.twitter.com/en/rules-and-policies/twitter-rules>
- Waldron, J. (2010). *Dignity and Defamation: The Visibility of Hate*. Abgerufen am 8. Juli 2019 von Harvard Law Review: <https://www.jstor.org/stable/40648494>
- Warner, W., & Hirschberg, J. (2012). *Detecting Hate Speech on the World Wide Web*. Association for Computational Linguistics.
- Windhager, M. (2016). *Hass-Kommentare im Netz - Rechtliche Aspekte*. Abgerufen am 13. Juni 2018 von Grünbuch Digitale Courage. Im Auftrag des Präsidenten des Bundesrates Mario Lindner:

[https://parlament.gv.at/ZUSD/PDF/Gruenbuch\\_Digitale\\_Courage\\_Republik\\_Oest\\_erreich\\_Bundesrat.pdf](https://parlament.gv.at/ZUSD/PDF/Gruenbuch_Digitale_Courage_Republik_Oest_erreich_Bundesrat.pdf)

Wissenschaftliche Dienste des Deutschen Bundestages. (2017). *Entwurf eines  
Netzwerkdurchsetzungsgesetzes: Vereinbarkeit mit der Meinungsfreiheit.*

Abgerufen am 13. Juli 2019 von

<https://www.bundestag.de/resource/blob/510514/eefb7cf92dee88ec74ce8e796e9bc25c/wd-10-037-17-pdf-data.pdf>

YouTube. (2019). *Community-Richtlinien*. Abgerufen am 17. Juli 2019 von

<https://www.youtube.com/intl/de/yt/about/policies/#community-guidelines>

Zankl, W. (2016). *ECG. E-Commerce-Gesetz: Kommentar. 2. Auflage*. Verlag  
Österreich.

ZARA - Zivilcourage & Anti-Rassismus-Arbeit. (2018). *Rassismus Report 2018.*

*Einzelfall-Bericht über rassistische Übergriffe und Strukturen in Österreich.*

Zeit Online. (2016). *Gericht bestätigt Strafe für Lutz Bachmann*. Abgerufen am 13. Juli  
2019 von <https://www.zeit.de/gesellschaft/zeitgeschehen/2016-11/pegida-lutz-bachmann-volksverhetzung-geldstrafe-urteil>

Zeit Online. (o.J.). *Netiquette*. Abgerufen am 19. Juli 2019 von

<https://www.zeit.de/administratives/2010-03/netiquette/seite-3>



## VIII. Anhang

### 8.1. Codebuch

#### **Codebook: Content Analysis of Policies against Harmful Online Communication**

Topic: Harmful online communication (HOC)

Ways of expression in online environments containing aggressive and destructive diction that violate *social norms and may harm the dignity or safety* of others.

Unit of Analysis = Document

Documents are drawn from the company's website. They are considered individual documents when they appear under a separate URL or when they are under the same URL but are treated with the same level of emphasis (i.e., heading is in same font size), so it is clear that for the company this is a separate policy aspect.

Only such documents are included in the sample that contain content on HOC (e.g., if the Terms of Service exist but do not contain anything regarding HOC, the document is not included in the sample)

#### **Coding Procedure**

##### **Step 1:**

Identify all content within the document that deals with HOC and mark the relevant content with yellow colour. If a paragraph comprises multiple sentences on different aspects (e.g., privacy and harmful content), we only count those sentences that deal with HOC content and not those that deal with something else (like privacy or copyright). This means, we use only full sentences, but not the full paragraph if it also contains sentences that do not contain HOC content.

We consider as **relevant content**

- any text that addresses the harming of individuals or groups by using harmful language and/or visuals and/or audio.
- This also includes passages on child protection if it addresses the harming of children through language/visuals/audio (e.g. mobbing, instigating suicide) and passages on pornography if it is implied that through the pornographic content someone was or will get hurt (e.g., child pornography, pornographic content that may be offensive for certain individuals/groups).

### **Step 2:**

Code the document and its content using the coding scheme sections I. and II.

NOTE for calculating Intercoeder Reliability: fill blanks with the number 99, since ReCal does not work when there are missing values. (see <http://dfreelon.org/utills/recalfront/>)

## **I. Form**

### **1. Document Identification Number (ID)**

*CountrycodeCompanynameDocumenttype*

For countrycode see # 4 COUNTRY

For document type see #7 TYPE.

If the company has two or more documents of the same type, we extend every document's Type-number with -1, -2 etc. If there is only one document of one Type we do not use any extension.

Example:

- there is only one document of Type = 1 for Facebook UK: UKfacebook1
- there are two documents of Type = 1 for Facebook UK: UKfacebook1-1, UKfacebook1-2

### **2. Name of the company (COMPANY)**

*Please note down the name of the company that published the document on its site.*

Open

### **3. Medium (MEDIUM)**

*Type of medium to which the document refers.*

- 1 – Web Portal Sites
- 2 – Blog Hosting Sites
- 3 – News Media Sites
- 4 – Social Network Sites
- 5 – Community Sites
- 6 – E-Commerce Sites
- 7 – Corporations' SNSs (Social Network Sites)
- 8 – Corporations' Websites
- 9 – Recommendation Portals

### **4. Country (COUNTRY)**

*Country in which the medium is published*

- 1 – Austria (AU)
- 2 – Germany (GE)
- 3 – USA (US)
- 4 – UK (UK)

### **5. Type of document (TYPE)**

*Companies may have different types of documents containing aspects regarding HOC. What type of document is this one? Please note that the actual names of the document may be different.*

- 1 – Terms of Use/Service, Conditions of Use, Services Agreement, Operation Policy (generally more formal/legally sounding document regarding terms how to use the site)
- 2 – Community Guideline, Community Standards, Netiquette, Mission, Rules (of the Road, for users, for using, of the game), Principle of Community (defines the way of interacting among community members and with community provider)
- 3 – Content Guideline (specifies how the content can or shall not be worded or look like)
- 4 – Reporting Guideline (informs on how to report abuse or violation of terms/rules)

## II Content

### 6. Addressing the interactions between users (INTERACT)

*Is the way users should interact with each other addressed in the part of the document in which HOC is covered? How to treat/interact with others has to be explicitly addressed.*

0 – no, how to treat/ interact with others is not mentioned

1 – yes, the text explicitly mentions how to treat/ interact with others (even if it is only one sentence)

Examples:

- „Respect Others - Please be polite to all the members of our Community, including other commenters, authors and the subjects of articles. Also, keep in mind that there are real people reading your comments.“
- „Wenn Du wütend oder schlecht drauf bist, lass Deine schlechte Laune nicht an anderen Usern aus!“
- „Damit sich jeder innerhalb der BILD-Community wohlfühlt, gelten auch hier bestimmte Umgangsformen. Nur gegenseitiger Respekt und Höflichkeit sorgen für eine rege Diskussion.“

The following variables (HANDLE 1 to 5 & 10 to 11) regard the different ways in which the company may handle HOC.

### 7. Delete without explanation (HANDLE1)

*Does the company state that it will delete prohibited content without providing an explanation? Code yes also when it is not explicitly stated that “no explanation” is given.*

0 – no

1 – yes

Examples:

- „Your comment may be removed at any time, if a Moderator feels it violates the Posting Guidelines or Code of Conduct.“
- „Die Moderatoren und Administratoren werden Beiträge, die gegen die Regeln verstoßen, sofort nach Kenntnisnahme löschen.“
- „Bei Nichtbeachtung der Nutzungsbedingungen behält die „Welt“ sich vor, einzelne Beiträge (ohne vorherige Abstimmung) zu löschen, zu bearbeiten, zu verschieben, zu schließen oder einzelnen Nutzern zeitweise oder dauerhaft den Zugang zu Leserkomentaren zu sperren.“

### **8. Delete or curtail with explanation or comment (HANDLE2)**

*Does the company state that it will delete prohibited content with providing an explanation or commentary why it acted that way? Here, the company must explicitly mention that it will provide an explanation or comment.*

0 – no

1 – yes

Examples:

- „Wir begründen unser Einschreiten meist durch kurze Anmerkungen und kennzeichnen, an welchen Stellen wir Kommentare gekürzt haben.“

### **9. Warning of user (HANDLE3)**

*Does the company state that it will warn the user who posted prohibited content?*

0 – no

1 – yes

Examples:

- „Die folgenden fünf Regeln gelten für alle Nutzer der Kommentarfunktion und führen bei Nichtbeachten zu einer Verwarnung und im Wiederholungsfall zu einer Sperrung des Nutzers.“

### **10. Delete or close user's account (HANDLE4)**

*Does the company state that it will exclude, ban or block the user or delete or close the user's account who posted prohibited content?*

*Es wird 1 – yes kodiert, wenn Maßnahmen genannt werden, die dazu führen, dass der User ausgeschlossen wird und nicht mehr teilnehmen kann. Dies kann delete, exclude, close, block oder ban sein, da alle diese Maßnahmen den Ausschluss des Users bewirken.*

0 – no

1 – yes, exclude, ban, block the user or delete/close the account

Examples:

- „Individuals who consistently or intentionally post these types of comments may lose their ability to comment and, if we deem necessary, be permanently excluded from the community.“
- „Die folgenden fünf Regeln gelten für alle Nutzer der Kommentarfunktion und führen bei Nichtbeachten zu einer Verwarnung und im Wiederholungsfall zu einer Sperrung des Nutzers.“

- „Bei Nichtbeachtung der Nutzungsbedingungen behält die „Welt“ sich vor, einzelne Beiträge (ohne vorherige Abstimmung) zu löschen, zu bearbeiten, zu verschieben, zu schließen oder einzelnen Nutzern zeitweise oder dauerhaft den Zugang zu Leserkomentaren zu sperren.“

### **11. Delete or close the whole discussion (HANDLE5)**

*Does the company state that it will delete or close the whole discussion if (usually multiple) users post prohibited content and thus violate the policy/ies?*

0 – no

1 – yes

### **12. Legal/judicial persecution (HANDLE10)**

*Does the company state that it will either sue or legally/judicially persecute users who posted prohibited content? This also includes when the company states that the user who posted prohibited content will be charged in case a third party sued the user.*

0 – no

1 – yes

### **13. Other (HANDLE11)**

*Does the company state any other way that it will handle prohibited content?*

0 – no

1 – yes: Open - please note down briefly in the Excel file (do not insert this variable when testing intercoder reliability with ReCal)

*The following variables (ACTION 1 to 8) regard the different ways in which the user may take action against HOC.*

### **14. Possibilities to mark the post (ACTION1)**

*Are users given the opportunity to take action against HOC by marking a post/comment right next to the post/comment? Does the company provide a “reporting icon”, “tick box” or the like to mark the content on the site itself?*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, there is a reporting icon or tick box on the site (e.g., standard on facebook), but it is not mentioned in the document

Example:

- „If you see a violation of these guidelines, report it to Facebook by clicking the "X" in the upper right corner of the comment in question, and marking the comment as spam or abuse.“

**15. Possibilities to notify the provider – unstandardized Email (ACTION2)**

*Are users given the opportunity to notify the company/ platform host about HOC by contacting them by sending them an email? This notification is done by the user in an unstandardized way, i.e. there is no template or form the user can use.*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, but it is not mentioned in the document

Example:

- “If you believe that any content on the Sites violates the Terms of Use, please notify the Sites by sending an email to [termsofuse@businessinsider.com](mailto:termsofuse@businessinsider.com).“

**16. Possibilities to notify the provider – unstandardized Telephone (ACTION3)**

*Are users given the opportunity to notify the company/ platform host about HOC by contacting them by phone call/hotline? This notification is done by the user in an unstandardized way, i.e. there is no template or form the user can use.*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, but it is not mentioned in the document

**17. Possibilities to notify the provider – unstandardized Postal Mail (ACTION4)**

*Are users given the opportunity to notify the company/ platform host about HOC by contacting them by postal mail/ writing a letter? This notification is done by the user in an unstandardized way, i.e. there is no template or form the user can use.*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, but it is not mentioned in the document

**18. Possibilities for users to notify the provider - standardized Template (ACTION5)**

*Are users given the opportunity to notify the company / platform host about HOC by using a specifically designed template, or is there a “reporting center” by the company that provides templates and instructions on how to report HOC. This template can be used to report prohibited content electronically or also to send the template via postal mail.*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, but it is not mentioned in the document

**19. Possibilities for users to notify an authority or government agency (ACTION6)**

*Are users given the opportunity to notify an authority or government agency about HOC. To do so, the company includes a link on the site to the authority/ agency where HOC can be reported.*

0 – no

1 – yes, and it is mentioned in the document

2 – yes, but it is not mentioned in the document

**20. Call to action to take personal action in the form of counter-speech (ACTION7)**

*Are users encouraged in the document to become active and counter HOC themselves by writing counter-speech? The company may even provide explicit instructions on how to write counter-speech or provide a link to a site that provides these instructions.*

0 – no

1 – yes

Example:

- „We’re a global community of people with diverse beliefs, opinions, and backgrounds, so please be respectful and keep hateful and incendiary comments off of Yahoo. Please read [tips for confronting hate speech](#) from the Anti-Defamation League.“

**21. Other possibility for users to take action (ACTION8)**

*Does the company state any other way that that user can take action?*

0 – no

1 – yes: Open - please note down briefly in the Excel file (do not insert this variable when testing intercoder reliability with ReCal)



### **III Detailed Content Analysis**

*The following variables (HOW1 to 4) will focus on how the users should behave according to the document.*

#### **22. Responsible for content (HOW1)**

*Are users being told that they are responsible for their submitted content?*

0 – no

1 – yes

Examples:

- “You agree that you are fully responsible for the content that you submit. You will promptly remove any content that you have posted should you discover that it violates these rules or that it is otherwise inappropriate.”
- “You are responsible for all Content that you upload, post, transmit or otherwise make available via the Service.”

#### **23. Responsible manner (HOW2)**

*Are users being told that they should behave in a responsible manner?*

0 – no

1 – yes

Examples:

- “When people share anything on Facebook, we expect that they will share it responsibly, including carefully choosing who will see that content.”
- “Seien Sie verantwortungsvoll, geben Sie sich nicht als jemand anderer aus, verwenden Sie nicht falsche Referenzen an und täuschen Sie nicht Fachkenntnisse vor, über die Sie nicht verfügen.”

#### **24. Respect (HOW3)**

*Are users being asked to treat each other with respect?*

0 – no

1 – yes

#### **25. Reporting of unwanted content (HOW4)**

*Are users explicitly being asked to report content that violates the policy?*

0 – no

1 – yes

*The following variables (FORBIDDEN1 to 6) regard prohibited content or behaviours according to the document.*

**26. Pornography (FORBIDDEN1)**

*Are the terms “pornography”, “nudity”, “obscene”, “sexually explicit”, “vulgar” or similar terms mentioned within the document?*

0 – no

1 – yes

**27. Hate speech (FORBIDDEN2)**

*Are the terms “hate speech”, “hateful”, “hate”, “hatred” or similar terms mentioned within the document?*

0 – no

1 – yes

**28. Illegal (FORBIDDEN3)**

*Are the terms “illegal”, “criminal behaviour”, “violence”, “unlawful”, “fraudulent”, “libellous” or similar terms mentioned within the document?*

0 – no

1 – yes

**29. Harassment (FORBIDDEN4)**

*Are the terms “harassment”, “bullying”, “defamatory”, “derogatory”, “offensive”, “degrading”, “intimidating”, “abusive”, “threatening” or similar terms mentioned within the document?*

0 – no

1 – yes

**30. Discrimination (FORBIDDEN5)**

*Are gender, race, class, religion, sexual orientation or ethnicity discriminations and similar terms mentioned within the document?*

0 – no

1 – yes

### **31. Privacy violations and third-party rights (FORBIDDEN6)**

*Are the terms “privacy violations” or similar terms mentioned within the document?*

0 – no

1 – yes

### **32. Child protection (CHILD)**

*Is child and youth protection or are measures for child and youth protection mentioned in the document? Does the document mention that it is prohibited to publish certain content to protect children and young people?*

0 – no

1 – yes

*The following variables (PURPOSE1 to 3) regard the purpose of the policies. They focus on the different reasons why some contents or behaviours are prohibited.*

### **33. Freedom of speech (PURPOSE1)**

*Is “freedom of speech” or are any similar terms mentioned within the document?*

0 – no

1 – yes

Examples:

- “Disqus enables online discussion communities, and in doing so, freedom of expression and identity are core values of the Service. There are a number of categories of content and behavior, however, that jeopardize the Service by posing risk to users, publishers or third party services utilizing the Disqus platform.”

### **34. Encouraging healthy debates (PURPOSE2)**

*Are some contents or behaviours prohibited in order to encourage a healthy debate or are users being encouraged to have objective and fair discussions?*

0 – no

1 – yes

Examples:

- “The Company wishes to encourage the broadest exchange of news and information to encourage stimulating, informative and above all intelligent debate.”

### **35. Rules of good conduct (PURPOSE3)**

*Does the document mention any ground rules of good conduct such as ethical behaviour, sophisticated communication or similar values?*

0 – no

1 – yes

Examples:

- “Allerdings gelten sowohl für das Forum als auch für die Postings und Weblogs einige Spielregeln, die beachtet werden müssen, damit Qualität, Niveau und rechtliche Sicherheit gewährleistet bleiben.”
- “Jeder soll zu unseren Themen seine Meinung abgeben können. Das muss sich aber auf einem vernünftigen Niveau bewegen und darf niemanden beleidigen oder sonstwie ernsthaft beeinträchtigen.”


### **36. Instructions (INSTRUCTION)**

*Does the document contain any detailed reporting instructions?*

0 – no

1 – yes

Examples:

- “Klicken Sie in der Antwort oder dem Kommentar auf das Missbrauch melden-Symbol . Wählen Sie den Missbrauchstyp aus, der den ausgewählten Inhalt am genauesten beschreibt. Tipp: Klicken Sie auf mehr, um eine detaillierte Beschreibung jedes Verstoßes zu lesen. Geben Sie Kommentare über die Meldung im Dialogfeld „Weitere Details“ ein. Klicken Sie auf Missbrauch melden.”
- “If you encounter member-generated content that you find objectionable, and if the tools are not sufficient to address the member who you believe is violating the community guidelines, you can alert AOL by clicking the Report Spam or Notify AOL button.”
- “If you encounter a blog that you believe violates our policies, please report it to us using the 'Report Abuse' link located at the top of each blog under the 'More' dropdown. If the blog owner has hidden this link, you can still report abuse in the Blogger Help Center.”

### 37. Reported content (REPORTED)

*Does the document mention what happens with the reported content? It does not include general possibilities of action in the eventuality of violations of the guidelines. It only includes whether or not the user learns what happens to his reported content.*

0 – no

1 – yes

## 8.2. Reliabilitätstest

### ReCal 0.1 Alpha for 2 Coders results for file "Daten für Reliabilitätstest.csv"

File size: 5461 bytes  
N columns: 32  
N variables: 16  
N coders per variable: 2

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha (nominal)	N Agreements	N Disagreements	N Cases	N Decisions
Variable 1 (cols 1 & 2)	100%	1	1	1	84	0	84	168
Variable 2 (cols 3 & 4)	100%	1	1	1	84	0	84	168
Variable 3 (cols 5 & 6)	100%	1	1	1	84	0	84	168
Variable 4 (cols 7 & 8)	100%	1	1	1	84	0	84	168
Variable 5 (cols 9 & 10)	100%	1	1	1	84	0	84	168
Variable 6 (cols 11 & 12)	100%	1	1	1	84	0	84	168
Variable 7 (cols 13 & 14)	100%	1	1	1	84	0	84	168
Variable 8 (cols 15 & 16)	100%	1	1	1	84	0	84	168
Variable 9 (cols 17 & 18)	100%	1	1	1	84	0	84	168
Variable 10 (cols 19 & 20)	100%	1	1	1	84	0	84	168
Variable 11 (cols 21 & 22)	100%	1	1	1	84	0	84	168
Variable 12 (cols 23 & 24)	100%	1	1	1	84	0	84	168
Variable 13 (cols 25 & 26)	100%	1	1	1	84	0	84	168
Variable 14 (cols 27 & 28)	100%	1	1	1	84	0	84	168
Variable 15 (cols 29 & 30)	100%	1	1	1	84	0	84	168
Variable 16 (cols 31 & 32)	100%	1	1	1	84	0	84	168

# Abstract Deutsch

Die vorliegende Arbeit erforscht anhand einer Inhaltsanalyse von Unternehmensrichtlinien aus Österreich, Deutschland, den Vereinigten Staaten und Großbritannien, wie Unternehmen mit Harmful Online Communication (HOC) umgehen. Hierfür werden zunächst die wichtigsten Begriffe erklärt und die häufigsten Akteure und ihre Motive vorgestellt. Außerdem werden die zentralen Faktoren genannt, die Hass im Netz begünstigen können. Darunter fallen unter anderem der „Nasty Effect“ von Anderson et al. (2014) und der „Online Disinhibition Effect“ von John Suler (2004). Darüber hinaus wird auch dargelegt, wo in den unterschiedlichen Ländern die Grenze zwischen der freien Meinungsäußerung und HOC gezogen wird. Ferner wird auch erklärt, inwiefern die Unternehmen eine soziale Verantwortung tragen und welche Maßnahmen sie treffen können, um gegen HOC vorzugehen.

Ziel dieser Forschungsarbeit war es herauszufinden, welche Inhalte laut der Unternehmensrichtlinien zu HOC zählen, welche Maßnahmen von den Unternehmen getroffen werden, um gegen HOC vorzugehen und inwiefern länderspezifische Unterschiede in Bezug auf HOC festgestellt werden können. Insgesamt wurden hierfür 419 Richtlinien von 152 Unternehmen analysiert. Es hat sich gezeigt, dass viele unterschiedliche Inhalte unter HOC fallen und deshalb auf den Online-Plattformen verboten werden. Hierzu zählen unter anderem obszöne, illegale, hassgefüllte, beleidigende, anstößige oder diskriminierende Äußerungen und Inhalte. Weiters konnte festgestellt werden, dass Unternehmen unterschiedliche Maßnahmen treffen, um gegen HOC vorzugehen. Unter anderem kommunizieren sie in ihren Richtlinien, wie sich die Userinnen und User auf ihren Plattformen verhalten sollen, weshalb gewisse Inhalte nicht toleriert werden können und welche Konsequenzen bei einem Verstoß gegen die Richtlinien drohen. Mit diesen proaktiven und restriktiven Maßnahmen wollen sie für ein faires, gesundes und freundliches Diskussionsklima sorgen. Letztendlich hat diese Forschungsarbeit ergeben, dass nur geringe länderspezifische Unterschiede bestehen, da die Richtlinien inhaltlich überwiegend ähnlich sind.

**Stichwörter:** Harmful Online Communication, Hate Speech, Hass im Netz, Trolle, Glaubenskrieger, Motive, Anonymität, Nasty Effect, Online Disinhibition Effect, Echokammern, Filterblasen, Recht auf freie Meinungsäußerung, Soziale Verantwortung, Lösungsansätze, Counter Speech.

## Abstract English

A content analysis of company policies from Austria, Germany, the United States and Great Britain has been conducted in this paper to explore how companies deal with Harmful Online Communication (HOC). For this purpose the most important terminologies as well as prominent players and their motives are specified. Furthermore the main factors that can encourage hatred on the Internet are introduced. These include the "Nasty Effect" by Anderson et al. (2014) and the "Online Disinhibition Effect" by John Suler (2004). The paper also points out where the lines are drawn between freedom of expression and HOC in the respective countries. In addition, it states to what extent companies do have a social responsibility and what measures they can implement to take action against HOC.

Aim and purpose of this research were to find out which contents are considered HOC according to the company policies, what measures are being taken by companies to take action against HOC and to what extent country-specific differences in HOC can be identified. A total of 419 policies of 152 companies have been analysed. It became apparent that many different kinds of contents can be classified as HOC, which is why they are prohibited on the online platforms. Among these are obscene, illegal, hateful, offensive, abusive or discriminatory comments or contents. It was also found out that companies take various measures to take action against HOC. They communicate rules of behaviour to indicate how users should act on their online platforms, why certain contents cannot be tolerated and what consequences may ensure if these policies are being violated. With these proactive and restrictive measures, organizations seek to provide a fair, healthy and friendly environment for discussions. Finally, this research has shown that there are only few country-specific differences, as the contents of policies tend to be very similar.

**Keywords:** Harmful Online Communication, Hate Speech, Trolls, Motives, Anonymity, Nasty Effect, Online Disinhibition Effect, Echo Chambers, Filter Bubbles, Freedom of Expression, Social Responsibility, Counter Speech.