

# **Masterarbeit**

Hate Speech in Zeiten von Social Media – Die  
Regulierung von sozialen Netzwerken zur Vermeidung  
von Hassreden

Von:

Lisa Berke  
mm171803

Begutachter:

FH-Prof. Priv.-Doz. Dr. Michael Litschka

Zweitbegutachter:

FH-Prof. Priv.-Doz. Dr. Andreas Gebesmair

St. Pölten, am 26.08.2019

# Ehrenwörtliche Erklärung

Ich versichere, dass

- ich diese Masterarbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- ich dieses Masterarbeitsthema bisher weder im Inland noch im Ausland einem Begutachter / einer Begutachterin zur Beurteilung oder in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Diese Arbeit stimmt mit der vom Begutachter beurteilten Arbeit überein.

.....  
Ort, Datum

.....  
Unterschrift

## **Zusammenfassung**

Diese Masterthese befasst sich mit der Thematik der Regulierung von sozialen Netzwerken zur Vermeidung von Hate Speech. Es soll dabei herausgefunden werden, welche Maßnahmen gesetzt werden müssen um zukünftig effektiv gegen Hassreden in den sozialen Netzwerken vorzugehen.

Diese Forschungsarbeit gliedert sich in zwei Teile, in einen theoretischen und einen empirischen Teil. Ersteres umfasst die den Themen, Social Media, Hate Speech sowie Regulierungen von sozialen Netzwerken. Abgeschlossen wird dieser Abschnitt mit einem Zwischenfazit. Daraufhin folgt der empirische Teil, der die Methodik, die Empirie sowie die Ergebnisse dieser Arbeit beschreibt.

Mittels der Literaturrecherche des theoretischen Teils dieser Forschungsarbeit konnte ermittelt werden, dass die sozialen Netzwerke heute eine große Rolle bei der Kommunikation sowie der Meinungsbildung der NutzerInnen spielen. Hate Speeches führen in den sozialen Netzwerken jedoch dazu, dass der öffentliche Diskurs beeinträchtigt wird. Daher ist es wichtig Regulierungen für Plattformunternehmen wie Facebook oder Twitter zu finden, die das Aufkommen von Hate Speeches reduzieren. Obwohl die EU, sowie Staaten innerhalb der EU, mittlerweile Gesetze zur Regulierung eingeführt haben, sind diese noch nicht ausreichend. Daher müssen zukünftig Maßnahmen gesetzt werden, um effektiv gegen Hass in den sozialen Netzwerken vorgehen zu können. Diese Erkenntnisse, die im theoretischen Teil ermittelt werden konnten, stellten die Basis für die empirische Untersuchung dar. Um am Ende dieser Arbeit die vier Forschungsfragen beantworten zu können, wurde die Methode der ExpertInneninterviews gewählt. Die empirische Untersuchung hat ergeben, dass eine enge Zusammenarbeit der Politik, Plattformen sowie Meldestellen notwendig ist. Es ist wichtig, dass auf die Strukturen der sozialen Netzwerke bei der Regulierung eingegangen wird, dass mehr Bildungsarbeit bei den NutzerInnen geleistet wird und dass die sozialen Netzwerke mehr menschliche Moderation zur tatsächlichen Erkennung von Hassreden anwenden. Ein zentrales Problem ist jedoch, dass der Begriff der Hate Speech gesellschaftlich keine einheitliche Definition aufweist.

## Abstract

This master thesis is about the regulation of social networks to prevent hate speech. It is about to find out which measures have to be taken to take action against hate speech in social networks effectively in the future.

The research paper is structured in two parts, a theoretical and an empirical part. The former one is about the topics social media, hate speech as well as the regulation of social networks. This part concludes with an interim conclusion. After that the empirical part follows, which describes the methodology, the empiricism and the results of this paper.

Through the literature research in the theoretical part of this paper it could be found out, that social networks play a huge part in the communication and the opinion making of users. However, hate speeches cause a restriction of public discourses in the internet. Due to that it is important to find an appropriate regulation for platform-companies like Facebook or Twitter which are able to reduce the number of hate speeches. Even though the EU as well as states of the EU already passed laws, these are not sufficient. That is why measures have to be implemented to take action against hate in the social networks. These perceptions of the theoretical part provide the base for the empirical research. To answer the four research questions at the end of this research paper, interviews with experts have been conducted. The empirical research found out, that a tight collaboration of politics, platform-companies and advice centres is necessary. When it is about the regulation of social networks it is important to respond to the structures these platform-companies have. Also it is necessary that more educational work will be carried out and that social networks use more human moderation for the recognition of hate speech. However, the main problem is that there is no consistent definition of hate speech.

# Inhaltsverzeichnis

1	Einleitung .....	1
1.1	Einführung in die Thematik .....	1
1.2	Problematik und Zielsetzung .....	1
1.3	Aufbau und Methodik der Arbeit .....	3
1.4	Forschungsstand .....	3
2	Die Sozialen Medien .....	9
2.1	Was sind Soziale Medien? .....	9
2.2	Der Wandel der Öffentlichkeit .....	11
2.3	Die Veränderung der Kommunikation .....	13
2.4	Der Einfluss auf die Meinungsbildung .....	15
2.5	Zusammenfassung .....	17
3	Hate Speech in den Sozialen Medien .....	19
3.1	Was ist Hate Speech? .....	19
3.2	Die Unterscheidung zwischen online/offline Hate Speech .....	24
3.3	Die Auswirkungen von Hate Speech .....	26
3.4	Zahlen und Fakten zu Hate Speech .....	28
3.5	Mediale Beispiele zu Hate Speech .....	33
3.5.1	FPÖ-Landesrat wegen Hassposting angeklagt .....	33
3.5.2	FPÖ-Politiker wegen Hassposting aus Partei geworfen .....	33
3.5.3	Hassposting gegen Neujahrssbaby .....	34
3.5.4	Eva-Glawischnig gegen Facebook .....	34
3.6	Zusammenfassung .....	35
4	Die Regulierung von sozialen Netzwerken .....	37
4.1	Haftung als Plattformbetreiber .....	37
4.1.1	Regulierungsformen .....	38

4.2	Der Verhaltenskodex der EU .....	40
4.2.1	Wie soziale Medien gegen Hass im Netz vorgehen .....	43
4.3	Nationale Gesetze in den EU-Staaten .....	45
4.3.1	Österreich: Aufhebung der Anonymität.....	45
4.3.2	Deutschland: Netzwerkdurchsetzungsgesetz.....	47
4.3.3	Lettland: „Law On Information Society Services“ .....	49
4.3.4	Großbritannien: Unabhängiges Aufsichtsorgan .....	49
4.4	Die Problematik mit der Einschränkung der Meinungsfreiheit .....	50
4.5	Zusammenfassung.....	52
5	Zwischenfazit & Forschungsfragen .....	53
6	Methodik .....	55
6.1	Forschungsdesign.....	55
6.2	Erläuterung der Erhebungsmethode .....	56
6.2.1	Leitfadengestütztes ExpertInneninterview .....	56
6.2.2	Auswahl der ExpertInnen .....	57
6.2.3	Leitfaden.....	59
6.3	Umsetzung der Erhebungsmethode .....	61
6.3.1	Pretest.....	61
6.3.2	Interviewablauf .....	61
6.3.3	Aufzeichnung und Transkription .....	64
6.4	Erläuterung der Auswertungsmethode .....	65
6.4.1	Gütekriterien.....	67
7	Empirische Analyseergebnisse .....	69
7.1	Kategoriesystem A: Die Problematik der Hate Speech in sozialen Netzwerken.....	70
7.2	Kategoriesystem B: Die derzeitige Situation der Regulierung von sozialen Netzwerken.....	73

7.3 Kategoriesystem C: Maßnahmen für eine effektive Regulierung von sozialen Netzwerken .....	78
7.4 Interpretation der Ergebnisse.....	84
7.5 Beantwortung der Forschungsfragen.....	86
7.5.1 Forschungsfrage 1.....	86
7.5.2 Forschungsfrage 2.....	87
7.5.3 Forschungsfrage 3.....	88
7.5.4 Forschungsfrage 4.....	89
7.6 Hypothesengenerierung.....	90
8 Conclusio .....	91
8.1 Limitation.....	91
8.2 Fazit .....	92
8.3 Ausblick.....	94

## Literaturverzeichnis

## Anhang

## Abbildungsverzeichnis

Abbildung 1: Erfahrungen mit „Hass-Postings“ - Eigene Darstellung nach ISA 2017, S. 54.....	29
Abbildung 2: Reaktionen auf "Hass-Postings" – Eigene Darstellung nach ISA 2017, S. 55.....	29
Abbildung 3: Gegen welche Gruppen und Institutionen richtet sich der Hass (Eckes et al. 2018, S. 20).....	31
Abbildung 4: Gemeldete rassistische Fälle – Eigene Darstellung nach ZARA 2018, S. 13.....	32
Abbildung 5: HassposterInnen verbreiten Rassismus hauptsächlich als Reaktion in Kommentaren – Eigene Darstellung nach ZARA 2018, S. 13 .....	32
Abbildung 6: Zahl der Benachrichtigungen der IT-Unternehmen – Eigene Darstellung nach Jourová 2016, S. 3 .....	41
Abbildung 7: % der gelöschten Inhalte aller Benachrichtigungen - Eigene Darstellung nach Jourová 2016, S. 4 .....	41
Abbildung 8: Rate of notifications reviewed within 24 hours since the launch of the Code of Conduct (Jourová 2019, S. 2).....	42
Abbildung 9: Removals per IT Company – Eigene Darstellung nach Jourova 2019, S. 3.....	43
Abbildung 10: Auswertungs-Tabelle – Eigene Darstellung .....	67

## **Tabellenverzeichnis**

Tabelle 1: Übersicht Forschungsstand - Eigene Darstellung .....	8
Tabelle 2: Formen symbolischer Abwertungen – Eigene Darstellung nach Sponholz 2018, S. 83.....	21
Tabelle 3: Übersicht Unterschied zwischen Troll und Glaubenskrieger - Eigene Darstellung nach Brodnig 2016, S. 75.....	23
Tabelle 4: Auflistung der ExpertInnen – Eigene Darstellung.....	59
Tabelle 5: Transkriptions-Regeln – Eigene Darstellung .....	64
Tabelle 6: Kategoriesysteme – Eigene Darstellung.....	69

## **Abkürzungsverzeichnis**

KI	Künstliche Intelligenz
AI	Artificial Intelligence
NetzDG	Netzwerkdurchsetzungsgesetz

# 1 Einleitung

## 1.1 Einführung in die Thematik

Immer wird bei sozialen Netzwerken auf Hasspostings oder Hasskommentare gestoßen. Diese sind häufig herabsetzend und haben das Ziel der Verunglimpfung bestimmter Personen oder Personengruppen. Facebook, Twitter und Co. versuchen gezielt gegen Hassreden vorzugehen, jedoch hat die Vergangenheit gezeigt, dass dies nicht so funktioniert wie es soll. Aus diesem Grund versuchen die EU-Mitgliedstaaten effektiver, mittels nationaler Gesetze, dagegen vorzugehen.

## 1.2 Problematik und Zielsetzung

Es war noch nie möglich so einfach, ohne Aufwand, schnell Informationen zu teilen und miteinander zu kommunizieren und diskutieren. Seit dem Aufkommen der modernen Online-Kommunikation, die durch das Web 2.0 ermöglicht wurde, wird von einer veränderten medialen Öffentlichkeit gesprochen. Facebook, Twitter oder YouTube als soziale Netzwerke sind heute in der täglichen Kommunikation eine Notwendigkeit. Besonders an diesen Plattformen ist, dass jeder die Möglichkeit hat aktiv und selbstbestimmt in das Kommunikationsgeschehen einzugreifen. Dabei kann persönlich zum Sender avanciert werden und sich ungefiltert und uneingeschränkt der ganzen Welt mitteilen. (vgl. Seifter, 2017, S. 54)

Die Online-Kommunikation hat jedoch nicht nur positive Seiten, sondern weist auch negative Seiten auf, wie etwa beleidigende Kommentare auf Webseiten von Zeitungen auf Facebook oder Twitter. (vgl. ebd., S. 55)

Faktoren, die Hasspostings innerhalb der Online-Kommunikation bewirken, sind hauptsächlich die Anonymität und die physische Distanz beziehungsweise Unsichtbarkeit. Diese Anonymität im Internet aufzuheben ist jedoch widersprüchlich. Zum einen wäre eine Registrierungspflicht bzw. ein Authentifizierungzwang mit einer Art Einschränkung der Meinungsfreiheit gleichzusetzen, da die Bereitschaft zur Partizipation wegen sozialer Ängste

(Verlust des Arbeitsplatzes, gesellschaftliche Isolation, etc.) bei kritischer Meinungsäußerung nachvollziehbar ist. Zum anderen konnte bis jetzt nicht erwiesen werden, dass das Internet durch seine Konstitution, vor allem durch die Anonymität, Hassreden verursacht. (vgl. ebd., S. 57)

Nicht zu leugnen ist jedoch, dass Hassreden auch im Internet existieren. Beispielsweise sind diese auf Facebook auf diversen Facebook-Seiten und -Gruppen zu finden. Plattformen wie Facebook sind wegen der Möglichkeit der Personalisierung des Kontos auch bekannt dafür, eine Verstärkung der eigenen Meinung zu bewirken. Dies passiert deshalb, da Facebook-Angebote gezielt von ähnlich gesinnten Personen aufgesucht werden. Dies führt dazu, dass kaum inhaltliche Abwechslung oder argumentative Ausgewogenheit des Kommunikationsraums vorhanden sind. Bei diesem Phänomen wird von sogenannten „Echokammern“ gesprochen. Durch diese kommen Teilöffentlichkeiten zu Stande. Dadurch kommt der Gesellschaft die gemeinsame Basis für einen politischen Diskurs abhanden, sodass eine kollektive Meinungs- und Willensbildung erschwert wird. Dabei kommt es zu Hass im Netz, die politische und gesetzliche Relevanz hervorruft. (vgl. ebd., S. 58)

Die zukünftige Regulation von Online-Medien bezüglich Hassreden, der Schutz von Minderheiten, die virale Verbreitung von Fake-News in den Sozialen Medien und den Kampf um das Urheberrecht haben hitzige Debatten in der Europäischen Kommission ausgelöst. Innerhalb der PolitikerInnen in Europa herrscht dahingehend Einigkeit, dass schnell Lösungen gefunden werden müssen, um diese Phänomene zu stoppen. (vgl. Rozgonyi, 2018, S. 18)

Im Mai 2016 haben die Unternehmen Facebook, Twitter, Youtube und Microsoft freiwillig einem Verhaltenskodex zugestimmt. Der Zweck dieses Kodex war es besser gegen Hassreden vorzugehen und die „notice and take down“ Regel effektiver zu machen. Die Plattformen haben sich dabei verpflichtet illegale Inhalte innerhalb von 24 Stunden zu löschen. Wird dies nicht eingehalten so können diese Plattformen mit einer Geldstrafe von bis zu 50 Millionen Euro belangt werden. Die erwartende Wirkung erreichte dieser Kodex jedoch nicht. Obwohl solche illegalen Inhalte meistens sofort entfernt wurden, wurden diese vor der Entfernung schon so oft weiterverbreitet, sodass es viele andere NutzerInnen ebenfalls posten können. Diese weiterverbreiteten Inhalte werden jedoch ignoriert. Es kann also

gesagt werden, dass die bisherigen Bemühungen soziale Netzwerke auf EU-Ebene zu regulieren gescheitert sind. (vgl. Niklewicz, 2017, S. 29f.)

Jedoch versuchen auch EU-Mitgliedsstaaten mit nationalen Gesetzen Hassreden in den sozialen Netzwerken zu verringern (vgl. Deutscher Bundestag, 2018, S. 1ff.)

Bisher stoßen jedoch vieler dieser Gesetze auf großen Widerstand, auf Grund der Angst die Meinungsfreiheit damit einzuschränken (vgl. Fox, 2018).

Daher stellt sich die Frage wie heute am effektivsten gegen Hassreden in den sozialen Medien vorgegangen werden kann. Die Leitfrage dieser Arbeit lautet also: „Wie können soziale Netzwerke zukünftig reguliert werden um Hate Speech zu vermeiden?“

### **1.3 Aufbau und Methodik der Arbeit**

Diese Arbeit besteht aus einem theoretischen und einem empirischen Teil. Der theoretische Teil beginnt mit einer Erläuterung von sozialen Medien und deren Auswirkungen auf unsere Gesellschaft. Darauf folgt ein Kapitel, dass sich mit Hate Speech beschäftigt. Abschließend wird auf die Regulierung von sozialen Netzwerken eingegangen. Es wird beschrieben, welche Ansätze es bisher gibt und welche geplant sind und warum befürchtet wird, dass dadurch die Meinungsfreiheit eingeschränkt wird.

Im empirischen Teil wird zuerst das Verfahren der Experteninterviews beschrieben. Anschließend werden die ausgewählten ExpertenInnen sowie die Ergebnisse der Befragung beschrieben.

### **1.4 Forschungsstand**

#### **Freiheit versus Regulierung im Internet**

Die Studie vom Deutschen Institut für Vertrauen und Sicherheit im Internet in Kooperation mit dem Institut für Demoskopie Allensbach, die im Jahr 2013 veröffentlicht wurde, beschäftigt sich mit Fragen wie, mit welchen Vorteilen und Gefahren die Bevölkerung und vor allem NutzerInnen das Internet verbinden, wie sicher sie sich im Netz fühlen, welche Erfahrungen sie dabei gemacht haben und wie sie versuchen sich vor Risiken zu schützen. Im Fokus dieser Studie stand die

Fragestellung, welcher Stellenwert vor dem Hintergrund dieser Bewertung und Erfahrung einer immer freien und eigenverantwortlichen Nutzung des Internets zugeschrieben und wieweit die Notwendigkeit gesehen wird, das Internet und dessen Inhalte stärker zu kontrollieren und zu regulieren. Daneben wurden auch Fragen zur Intensität und zu den Schwerpunkten der Internet-Nutzung gestellt. (vgl. DIVSI, 2013, S. 5)

Bei dieser Studie wurden 1.487 mündlich-persönliche Interviews mit Personen ab 16 Jahren in der Bundesrepublik Deutschland durchgeführt. Die Befragungen fanden zwischen dem 1. und 14. Oktober 2013 statt. (vgl. ebd.)

Die Ergebnisse dieser Studie haben aufgezeigt, dass Internet-NutzerInnen vor allem viele Vorteile mit dem Netz assoziieren. Daneben werden jedoch auch zahlreiche Risiken damit verbunden. Die wahrgenommen Gefahren des Internet hängen jedoch nicht mit eigenen negativen Ergebnissen zusammen und sind unabhängig von der Nutzung eines Einzelnen. Des Weiteren ist die Mehrheit der Befragten der Meinung, dass Internet-NutzerInnen zu sorglos mit den persönlichen Daten und den damit verbundenen Gefahren umgehen und zu viel von sich preisgeben. (vgl. ebd., S.53)

Wenn es um den Schutz vor Risiken im Internet geht sehen BürgerInnen den Staat gefordert. Die Mehrheit ist der Meinung, dass dieser dazu verpflichtet ist, den Plattformunternehmen gewisse Regeln vorzugeben und die Einhaltung dieser zu kontrollieren. Einigkeit besteht auch darüber, dass Maßnahmen auf nationaler Ebene zum Schutz von NutzerInnen nur wenig effektiv sein können. (vgl.ebd.)

Bezüglich der aktuellen staatlichen Einflussnahmen und Kontrolle von Internet-Seiten-Anbieter herrscht häufig Unklarheit und wird eher als zu gering eingeschätzt. Bei der Frage nach einer möglichlicherweise zu starken Kontrolle durch den Staat wird dies von etwa der Hälfte dahingehend als sehr kritisch betrachtet, als das die Freiheit im Internet potenziell durch die staatliche Überwachung bedroht werden könnte. Die Befragten sind jedoch der Meinung, dass eine noch größere Gefahr von den Unternehmen selber ausgehe, die nicht mehr alles, sondern zu meist nur mehr auf NutzerInnen-Profile zugeschnittene Inhalte zeigen. (vgl. ebd.)

## **Landesanstalt für Medien NRW: Hate Speech Ergebnisbericht**

Bei der Studie der Landesanstalt für Medien NRW wurden 1.008 Befragte im Alter ab 14 Jahren, die das Internet privat nutzen, online zum Thema Hassreden beziehungsweise Hasskommentare befragt. Dabei wurden die Befragten unter anderem zu ihrer Wahrnehmung von Hassreden beziehungsweise Hasskommentaren im Internet befragt. Die Mehrheit, etwas 78%, gaben dabei an, Hassreden im Internet schon einmal gesehen zu haben, beispielsweise auf Webseiten, in Blogs, sozialen Netzwerken oder in Internetforen. Die Altersgruppe der 14- bis 24-Jährigen ist schon weitaus häufiger als der Durchschnitt Hassreden im Internet begegnet. Bezüglich des Verhaltens bei der Wahrnehmung von Hassreden gaben 37% an, dass diese sich dadurch näher mit diesem Thema befasst haben, in dem diese zum Beispiel den Verlauf der Kommentare verfolgten, das Profil des Verfassers/der Verfasserin aufriefen oder mit anderen darüber gesprochen haben. Vergleichsweise dazu gaben 26% an, VerfasserInnen einer Hassrede schon einmal gemeldet zu haben. Nur eine Minderheit der Befragten hat angegeben, dass diese schon einmal auf einen Hasskommentar geantwortet haben um diesen zu unterstützen, selbst einen Hasskommentar geschrieben haben oder aufgrund eines Hasskommentars Anzeige bei der Polizei erstattet haben. InternetnutzerInnen unter 45 Jahren gaben viel häufiger an als InternetnutzerInnen über 45 Jahren, sich schon einmal näher mit Hasskommentaren beschäftigt zu haben, einen Hasskommentar beziehungsweise VerfasserInnen bei der entsprechenden Plattform gemeldet oder auf einen Hasskommentar geantwortet zu haben um diesen zu kritisieren. Zusätzlich dazu wurde danach gefragt, warum die Befragten sich mit Hasskommentaren im Internet beschäftigen. Die Mehrheit antwortete dabei, dass diese sich damit beschäftigen, weil es sich bei diesen um ein relevantes Thema handelt oder weil sie diese entsetzen. Ungefähr jeder Zweite, der schon einmal auf Hasskommentare reagiert hat, beschäftigt sich mit diesen, da diese etwas über die Meinung Anderer erfahren wollen und um diese gegebenenfalls zu melden. Vor allem jüngere NutzerInnen unter 25 Jahren beschäftigen sich sehr häufig mit Hasskommentaren, wenn es sich dabei um ein persönlich relevantes Thema handelt oder weil diese Themen interessant beziehungsweise unterhaltsam sind.

Bezüglich der Bewertung von Hasskommentaren im Internet haben 92% der Befragten angegeben, dass anonyme Hasskommentare feige sind. Der Großteil der Befragten gaben auch an, dass Hasskommentare sie wütend machen und es Zeitverschwendungen ist, sich mit diesen zu beschäftigen. Nur 19% der Befragten haben Verständnis für solche Hasskommentare im Internet. Zum Schluss wurde noch nach der Häufigkeit von im Internet gesehenen Hasskommentaren gefragt. Fast jeder zweite Befragte hat hierbei den Eindruck, dass durchschnittlichen InternetnutzerInnen Hasskommentare häufig oder sogar sehr häufig unterkommen. So gut wie niemand ist der Meinung, dass InternetnutzerInnen noch nie Hasskommentare im Internet begegnet sind. (vgl. Forsa., 2018, S. 1ff.)

### **#Hass im Netz: Der schleichende Angriff auf unsere Demokratie**

Im Rahmen einer hessischen Studie wurden 1.243 BürgerInnen in Hessen ab 18 Jahren zu deren Erfahrungen und Meinungen bezüglich Hassreden im Internet online befragt. Dabei kam heraus, dass über ein Drittel der Befragten im Internet bereits beleidigt worden sind. 13% haben sogar angegeben, dass ihnen im Netz bereits mit Gewalt gedroht worden sei. Diese haben ebenfalls angegeben, dass ihre Familien auch bereits beleidigt oder bedroht worden sind. Auffallend ist dabei, dass vor allem jüngere Menschen und Menschen aus Einwandererfamilien einen signifikant hohen Wert aufweisen. Bei den 18- bis 24-Jährigen haben 69% der Befragten angegeben, dass diese schon einmal online beleidigt wurden. Bei den befragten Personen aus Einwanderfamilien haben 21% berichtet, dass ihnen schon einmal körperliche Gewalt im Internet angedroht wurde. 41% der 18- bis 24-Jährigen Personen aus Hessen wurden im Internet bereits sexuell belästigt. Diese Belästigungen im Internet hatten für die meisten der Befragten eindeutige Konsequenzen. Bei den 18- bis 24-Jährigen gaben 83% an, negative Folgen erlebt zu haben. Zum Teil wurden dabei detailliert psychische und körperliche Beschwerden beschrieben. Durch diese Befragung wurde vor allem deutlich, dass besonders Jugendliche von Hass im Netz betroffen sind. Hierbei klagen 30% der 18- bis 24-Jährigen über Depressionen als Folge von Belästigungen im Netz. Die Mehrheit der Befragten haben hier auch angegeben, dass sie bereits aggressive und abwertende Botschaften gegen mindestens eine der abgefragten Gruppen oder Institutionen gesehen haben. Vor allem berichten die Befragten über Hass

gegenüber geflüchteten Menschen, Menschen mit Migrationshintergrund sowie Muslimas und Muslime. Jeder fünfte Befragte hat sogar angegeben „sehr oft“ abwertende oder aggressive Aussagen gegen geflüchtete Menschen im Internet gesehen zu haben. Es wurde ebenfalls abgefragt, auf welcher Plattform die Befragten am meisten Hasskommentare gelesen haben. Hierbei haben die meisten angegeben Hasskommentare auf Facebook, in Foren, auf kommentierbaren Blogs, auf Twitter sowie bei Instagram gesehen zu haben. (vgl. Eckes et al., 2018, S. S. 4ff.)

Bei der Frage nach Toleranz antwortete die Mehrheit damit, dass Hasskommentare, egal ob legal oder illegal, nicht toleriert werden sollten. Nur 16% sind der Meinung, dass Hasskommentare akzeptiert werden müssen. Diese fordern ebenfalls, dass mehr gegen Hass im Netz unternommen werden sollte. Vor allem verlangen sie ein stärkeres Engagement der hessischen Landesregierung. (vgl. ebd. S. 8)

Autor	Jahr	Titel	Methode	Fazit
DIVSI & Institut für Demoskopie Allensbach	2013	Freiheit versus Regulierung im Internet	Face-to-Face Interviews	Die Befragten sehen viele Vorteile in Nutzung des Internets, kennen jedoch auch die Gefahren und Risiken. Vor allem sind sie der Meinung, dass mit persönlichen Daten zu sorglos umgegangen wird. Daher fordern diese Regulierungen von Internet-Plattformen.
Landesanstalt für Medien NRW	2018	Hate Speech: Ergebnisbericht	Online-Befragung	Die Mehrheit der Befragten haben schon einmal Hassreden im Internet gesehen. 37% haben sich dabei schon einmal näher mit Hasskommentaren befasst, weil sie es für ein relevantes Thema hielten oder davon entsetzt waren. Bei der Frage nach der persönlichen

				Meinung bezüglich Hassreden gaben die meisten an, dass sie diese feige finden und dass Hasskommentare sie wütend machen.
Eckes Christine, Fernholz Tobiasl, Geschke Daniel, Klaßen Anja, Quent MAtthias	2018	#Hass im Netz: Der schleichende Angriff auf unsere Demokratie	Online-Befragung	Im Rahmen der Studie kam heraus, dass mehr als ein Drittel der Befragten bereits mit Hass im Netz zu tun hatten. Dabei sind vor allem Jugendliche beziehungsweise junge Erwachsene im Alter von 18 bis 24 Jahren betroffen, die folglich auch mit Depressionen zu kämpfen hatten. Der Hass im Netz richtet sich meistens gegen Flüchtlinge, Menschen mit Migrationshintergrund oder Muslimas und Muslime. Gelesen wurden die meisten Hasskommentare auf Facebook, in Foren oder auf kommentierbaren Blogs. Wenn es um die Toleranz geht, sind die meisten Befragten der Meinung, dass diese nicht toleriert werden sollten und mehr dagegen unternommen werden muss.

Tabelle 1: Übersicht Forschungsstand - Eigene Darstellung

Zusammengefasst kann also gesagt werden, dass es schon Studien zum Thema Regulierung von Internet-Plattformen sowie zu Hate Speech gibt. Anzumerken ist hier jedoch, dass im Rahmen der Recherche keine Studie gefunden wurde, bei der es um die Regulierung von sozialen Netzwerken zur Vermeidung von Hate-Speeches geht. Daher beschäftigt sich diese Arbeit mit diesem Bereich.

## 2 Die Sozialen Medien

In diesem Kapitel wird das Thema von Social Media genauer beleuchtet. Bevor die Auswirkungen dieser Plattformen auf die Gesellschaft beschrieben werden, wird der Begriff genauer erläutert. Daraufhin wird der Wandel der Öffentlichkeit sowie die Veränderung der Kommunikation thematisiert. Abschließend wird der Einfluss auf die Meinungsbildung erläutert.

### 2.1 Was sind Soziale Medien?

Laut Kaplan und Haenlein (2010) sind soziale Medien eine Gruppe internetbasierter Applikationen die auf der Technologie des Web 2.0 aufbauen und die die Kreation, sowie den Austausch von user-generated-content erlauben.

Soziale Medien werden zudem als digitale Medien definiert, die die menschliche Kommunikation mithilfe von algorithmisch vermittelten Social-Networking-Plattformen im Internet bereitstellen und unterstützen. Jedes soziale Medium schafft durch seine Programmierung den unternehmerischen Zweck sowie der damit verbundenen Vorstellung der Zielgruppe genaue Regeln und Formen der Kommunikation. (vgl. Schrape & Siri, 2019, S. 1053)

Soziale Medien schaffen die Möglichkeit unterschiedliche Informationen in Form von Bild, Text, Bewegtbild oder Sprache zwischen Individuen, die in irgendeiner Art zu einander in Beziehung stehen, unabhängig von Ort, Raum und Zeit zu speichern, verarbeiten und zu übermitteln. Dadurch entsteht eine Verbindung und Vernetzung von diesen Individuen zu einer Gruppe, wobei es für jedes Gruppenmitglied möglich ist den Austauschprozess aktiv zu gestalten. Ein soziales Medium muss jedoch nicht alle genannten Funktionen zur gleichen Zeit aufzeigen. Der Aspekt, der ein soziales Medium charakterisiert, ist hauptsächlich die Vernetzung der Individuen, die nicht dauerhaft sein muss, sondern auch zeitlich eingeschränkt werden kann. Häufig wird auch der Begriff „Soziale Netzwerke“ verwendet, der sich zu dem Begriff Social Media darin unterscheidet, dass dies eine Anwendungsklasse sozialer Medien ist. (vgl. Mack & Vilberger, 2016, S. 16f.)

Auf Grund der charakteristischen Netzwerke- und viralen Effekte haben sich diese Netzwerke räumlich und zeitlich mit einer großen Dynamik verbreitet. Neben der

Bereicherung des Medienangebots wurde dadurch auch die soziale Interaktion erweitert. Des Weiteren wurden Strukturen der Wertschöpfung verändert sowie neue Geschäftsmodelle entwickelt. Die Betreiber solcher Social-Media-Plattformen wurden über die Jahre zu bedeutsamen wirtschaftlichen und politischen AkteureInnen (vgl. Kolo, 2016, S. 1)

Global gesehen lassen sich derzeit sechs Dienste unterscheiden, die den Markt dominieren. Zum Ersten ist das, dass im Jahr 2004 gegründete Portal „Facebook“. Im Jahr 2018 verfügte das Unternehmen über zwei Milliarden aktive NutzerInnen. Mittlerweile dient das soziale Netzwerk den jüngeren UserInnen als primäre Quelle tagesaktueller Nachrichten. „WhatsApp“ zählt mit 1,2 Milliarden NutzerInnen im Jahr 2017 ebenfalls zu einem dominierenden Unternehmen in diesem Markt. Dieses soziale Medium wurde 2014 von Facebook aufgekauft und dient der nicht-öffentlichen persönlichen Kommunikation. Ebenfalls von Facebook aufgekauft wurde das soziale Netzwerke „Instagram“ im Jahr 2012. Diese Plattform ermöglicht das Verbreiten und Teilen von Bildern und Videos und wurde 2017 von über 800 Millionen Menschen verwendet. Auch „Snapchat“ ist sehr beliebt unter den jüngeren Social-Media-UserInnen. Seit 2011 ermöglicht dieses soziale Netzwerk das Versenden und Empfangen von Bildern und Videos, die sich nach kurzer Zeit selbst löschen. Vor allem in der Altersgruppe der 14- bis 29-Jährigen hat es sich als Alltagsmedium durchgesetzt. Im Jahr 2018 wurde „Snapchat“ täglich von 187 Millionen Smartphone-BesitzerInnen verwendet. Des Weiteren ist auch „Twitter“ eines der beliebtesten Social-Media-Plattformen. Der Microblogging-Dienst wurde 2006 gegründet und dient der Verbreitung telegrammarteriger Kurznachrichten. Im Jahr 2017 wurde der Dienst von 330 Millionen NutzerInnen verwendet. Das sechste dominierende Unternehmen am Social-Media-Markt ist das im Jahr 2005 gegründete und von Google akquirierte Videoportal „YouTube“. Dieses Netzwerk dient hauptsächlich der Verbreitung von nutzergenerierten und kommerziellen Videos. (vgl. Schrape & Siri, 2019, S. 1053f.)

All diese eben genannten Dienste sind in umfassende soziotechnische Ökosysteme eingelassen und bilden eine neuartige Form von einer semi-privater beziehungsweise persönlicher Öffentlichkeit. Diese prozessieren nebeneinander

und vereinfachen das Leben ihrer NutzerInnen durch viele zeitliche, sachliche und soziale Strukturierungsleistungen. (vgl. ebd., S. 1054)

## 2.2 Der Wandel der Öffentlichkeit

Der derzeitige Strukturwandel der Öffentlichkeit könnte als der Weg von der klassischen massenmedialen Welt der zweiten Hälfte des 20. Jahrhunderts in die Ära der digitalen Netzwerkmedien beschrieben werden. (vgl. Krüger, 2018, S. 13) Heute versuchen massenmediale Organisationen das Internet so gut es geht zu nutzen um auch dort Berichte, Reportagen oder Sonstiges zu veröffentlichen. Sie nutzen die unterschiedlichen Plattformen um dort gesellschaftsrelevante Informationen zu verbreiten. Hierbei müssen sie sich jedoch – wie auch offline – an ethische Standards des Journalismus halten. Das Internet erlaubt es aber eben auch normalen NutzerInnen Informationen zu verbreiten und unter gewissen Umständen auch das Massenpublikum zu erreichen. Teilweise wird heute dem Zeitalter der „BürgerjournalistInnen“ gesprochen, in dem alle BürgerInnen dank der Digitalkamera und dem Internetzugang über Ereignisse berichten können. Dies ist jedoch zu verneinen, da soziale Medien uns nicht automatisch zu JournalistInnen machen. Jedoch verliert der Journalismus sein Monopol auf die Aufbereitung und Veröffentlichung von Informationen. Die Medienöffentlichkeit beruht dank des Internets und den sozialen Medien nicht mehr nur auf journalistischer Vermittlungsleistung, sondern vielmehr auch auf partizipativer und technischer Vermittlung. Die Medienöffentlichkeit wird also von nicht-journalistischen NutzerInnen erweitert. (vgl. Schmidt, 2018, S. 49f.)

Im Unterschied zu der klassischen Medienöffentlichkeit werden bei der persönlichen Öffentlichkeit Informationen nach Kriterien der persönlichen Relevanz ausgewählt sowie aufbereitet. Das heißt also, dass ich auf meinem Blog, auf meiner Facebook-Seite oder meinem YouTube-Kanal das veröffentlichen kann, was ich erlebe, denke und was ich als wichtig empfinde. Bei der Aufbereitung sowie Präsentation von Themen folgt die persönliche Öffentlichkeit zudem andere Kriterien. Hier wird meist geschriebene Sprache vorgefunden, die sehr dem gesprochenen Wort ähnelt sowie umgangssprachliche Wendungen, Slang und eigene Abkürzungen und Emoticons. Wie im Journalismus herrscht

auch hier ein grundsätzlicher Anspruch auf Echtheit von Informationen, jedoch steht vielmehr die Wahrhaftigkeit im Vordergrund. Es wird von NutzerInnen erwartet sich weder zu verstehen noch sich für jemand anderen auszugeben. (vgl. ebd., S. 50f.)

Da die Fülle an Informationen im Internet kontinuierlich zunehmen, werden heute technische Hilfsmittel benötigt um diese erschließen zu können. Technische Infrastrukturen wie Softwareprogramme, Plattformen und Dienste des Internets und soziale Medien bringen dafür zusätzliche Mechanismen und Kriterien der Vermittlung mit ein. Diese eingesetzten Dienste und Programme erfassen, systematisieren, kategorisieren, katalogisieren und archivieren also die Informationswelten im Internet automatisch und teilweise in Echtzeit. So entstehen also für NutzerInnen unsichtbare Datenbanken. Durch die Suchanfrage und Verwendung bestimmter Schlagworte entscheiden NutzerInnen eines Informationsdienstes nun welche Themen aus den Informationen ausgewählt und dargestellt werden. Algorithmen bestimmten dabei, wie Informationen zu einem Thema oder einer Suchanfrage sortiert und dargestellt werden. Man merkt also ganz anders als in der Medien- oder der persönlichen Öffentlichkeit scheint die technische Vermittlung anderen und vor allem neutralen und objektiven Kriterien zu folgen, nach welchen sie Informationen freigeben. Beispielsweise wird bei einer Suchanfrage auf Google eine Seite weiter vorne platziert sein, je mehr andere Seite auf sie verlinken. Auch in den sozialen Medien kann eine Suchanfrage eingegeben werden. Allerdings sind hier andere Merkmale entscheidend ob und wie NutzerInnen Informationen angezeigt werden, nämlich die Aktualität und soziale Nähe. Bei dem Letzteren wird dies dadurch ermittelt, ob eine Person, ein Unternehmen, ein Medienangebot, eine Partei oder Ähnliches zu den Kontakten von NutzerInnen gehören. Ist dies der Fall, so werden mir Neuigkeiten auch angezeigt. Wenn ein solcher Kontakt nicht besteht, bekommen NutzerInnen keine Updates zu sehen, es sei denn, andere Kontakte leiten diese weiter, retweeten oder teilen diese und lenken sie auf den der NutzerInnen. Die Informationen aus dem Netzwerk von UserInnen werden, zusammen mit automatisch generierten Meldungen der Plattform, nach Aktualität sortiert sowie angezeigt und diese so, dass die neuesten Informationen als erstes angezeigt werden. (vgl. ebd., S.51ff.)

Zusammen mit dem anderen Radar ergibt sich also so eine besondere Form von Öffentlichkeit, die soziale Medien erzeugen. Es ist hier nicht der Fall, dass Informationen einen Einzelnen als Paket von Nachrichten zu einer bestimmten Zeit und nach journalistischen Kriterien ausgewählt, erreichen. Vielmehr ist es hier so, dass ein ständig aktualisierten und personalisierten Strom von Neuigkeiten , der dadurch bestimmt wird, welche Quellen zuvor diesem Strom zugeführt wurden, vorgefunden wird. Die Neuigkeiten von NutzerInnen beeinflussen auch die Ströme von Freunden und Bekannten in den sozialen Medien. Es ist also erkennbar, dass soziale Medien Öffentlichkeiten nicht nur verändert, sondern auch vernetzt haben. (vgl. ebd., S. 53f.)

## **2.3 Die Veränderung der Kommunikation**

22. Juli 2016, ein 18-jähriger Schüler erschoss in der bayrischen Hauptstadt München neun Menschen bevor er sich selbst tötete. Zu diesem Zeitpunkt änderten viele Fernsehsender ihr Programm und schalteten eine Sondersendung um über diesen Vorfall zu berichten. Für die meisten Menschen war jedoch das Internet eine genauso wichtige Anlaufstelle. Dort berichtete die Münchner Polizei nämlich laufend auf Facebook und Twitter über die aktuelle Situation vor Ort. Außerdem kursierten parallel dazu auf diesen Social-Media-Plattformen die ersten Augenzeugen-Videos. Besorgte Menschen meldeten sich per WhatsApp bei Familien und Freunden um sicher zu gehen, dass diese wohlauft sind. Außerdem aktivierte Facebook einen „Safety Check“, mit dem sich die Leute aus München und der Umgebung als „in Sicherheit“ markieren konnten. Auf Twitter versuchten rechtspopulistische PolitikerInnen jedoch nur mit fragwürdigen Postings Kapital aus der Tragödie zu schlagen. Im Gegensatz dazu engagierten sich BürgerInnen mit dem Hashtag #offenetür den Leuten Unterschlupf zu bieten, die durch dieses Drama in der Stadt gestrandet waren. (vgl. Schmidt, 2018, S. 7f.)

Anhand dieses Beispiels ist also erkennbar, dass sich das Gefüge der öffentlichen Kommunikation in den vergangenen Jahren deutlich verändert hat. Noch nie war es möglich so einfach, kostengünstig und schnell Informationen miteinander zu teilen, zu diskutieren oder sich gemeinsam mit anderen zu organisieren. Deshalb wird heute, wie schon erwähnt, von einer veränderten medialen Öffentlichkeit im

weiteren Sinne, die durch die moderne Online-Kommunikation durch das Web 2.0. ermöglicht wurde, gesprochen. Seit diesem Zeitpunkt ist die tägliche Kommunikation ohne die Sozialen Medien Facebook, Twitter oder YouTube nicht mehr vorzustellen. Eine entscheidende Eigenschaft von diesen Plattformen ist, dass es für jeden möglich ist aktiv und selbstbestimmt an dem Kommunikationsgeschehen teilzunehmen und einzutreten. Hierbei wird jeder persönlich zum Sender und kann sich potentiell ungefiltert und schrankenlos der ganzen Welt mitteilen. Im Vergleich zu den traditionellen Massenmedien produziert der Sender selbst Informationen und ist aus diesem Grund nicht mehr in der passiven Rolle der EmpfängerInnen beziehungsweise der KonsumentInnen. Wichtig zu erwähnen ist, dass in der Online-Kommunikation jeder gleichberechtigt ist, unabhängig von der Herkunft, Schicht und dem Geschlecht. (vgl. Seifter, 2017, S. 54ff.)

Neben der persönlichen Kommunikation hat sich also auch die Kommunikation von Medien verändert. Journalistische Inhalte werden heute nicht mehr nur durch die klassischen Medien in die persönliche Öffentlichkeit gebracht, sondern auch über die sozialen Medien. Dies geschieht entweder durch die eigenen Social-Media-Profile der Medienunternehmen oder durch die NutzerInnen, die bestimmte Artikel oder Beiträge verlinken und empfehlen. (vgl. Schmidt, 2018, S. 54f.)

Durch die Möglichkeit der problemlosen Veröffentlichung und Verbreitung von Informationen auf den sozialen Medien kommt es auch häufig vor, dass falsche oder missverständliche Informationen verbreitet werden. Die Gründe für die Publikation wie beispielsweise „Fake News“ sind unterschiedlich. Erwähnenswert ist jedoch, dass viele Menschen Informationen nicht hinterfragen, wenn diese in ihr Weltbild passen. Da die sozialen Medien die Meinungsbildung beeinflussen, hat Facebook 2016 beschlossen, stärker gegen irreführende Informationen vorzugehen. (vgl. ebd., S. 56ff.)

## 2.4 Der Einfluss auf die Meinungsbildung

Die sozialen Medien erstellen selber keine eigenen Inhalte. Jedoch hat ihre technische Gestalt einen großen Einfluss darauf, wie Informationen die Gesellschaft erreichen und wie diese sich darauf aufbauend eine Meinung bildet. Für viele NutzerInnen steigern die sozialen Medien die Vielfalt verfügbarer Informationen. Allerdings wurden Entwicklungen beobachtet, die in Filterblasen oder Echokammern enden können. Dadurch kommt es oft zu den Phänomenen wie „hate speech“ oder „social bots“, die den um gegenseitige Verständigung und Aufrichtigkeit bemühten Meinungsaustausch gefährden, den eine demokratische Gesellschaft benötigt. (vgl. Schmidt, 2018, S. 61)

Bevor hier nun genauer darauf eingegangen wird, müssen zuvor die dahinterstehenden Algorithmen dieser Plattformen erklärt werden, die im Kapitel 2.2. schon kurz erwähnt wurden.

Bei einem Algorithmus handelt es sich um eine Reihe von schrittweisen Befehlen, die durchgeführt werden sollen, um aus einer Eingabe eine Ausgabe zu bewirken um damit ein Problem zu lösen (vgl. Mahnke, 2015, S. 34f.). Diese Algorithmen werden heute zu den unterschiedlichsten Zwecken wie Automatisierung, Mustererkennung, Risikoanalyse oder für die Kontrolle in vielen Gesellschaftsbereichen verwendet (vgl. Humborg & Nguyen, 2018, S. 21). In der Medienbranche verwenden Unternehmen wie Netflix oder Buzzfeed dies zur Analyse von Nutzungsmuster ihrer NutzerInnen um prognostizieren zu können, welche Inhalte auf dem Werbe- und Publikumsmarkt gefragt sind (vgl. Heise, 2016, S. 202). Auch das soziale Netzwerk Facebook verwendet solche Algorithmen. Im Gegensatz zu journalistischen Redaktionen, die eine Reihe von Nachrichten werten, heranziehen um die Relevanz eines Themas zu bemessen, befolgt Facebook jedoch keine auf erkenntnistheoretischer Basis vordefinierte Kriterien. (vgl. Humborg & Nguyen, 2018, S. 21) Lischka und Stöcker (2017) sprechen hierbei von einer „komplexen Wechselwirkung zwischen NutzerInnenverhalten und algorithmischen Systemen“. Hierbei bereiten Plattformen eine vorselektierte Auswahl an Inhalten vor und lassen NutzerInnen mit diesen interagieren. Der Algorithmus merkt sich dabei welche Inhalte NutzerInnen anklicken gefallen, teilen oder auch ignorieren. All diese Daten

werden daraufhin ausgewertet und für den nächsten Selektionsprozess aufbereitet und verwendet. (vgl. ebd., S. 19ff.) Auch Faktoren wie das Surfverhalten, die Profilinformationen oder die Interaktion mit anderen NutzerInnen spielen hier eine Rolle. Für Facebook oder auch für Google ist die Personalisierung, also dass die Inhalte an die Präferenzen der einzelnen Personen angepasst werden, das oberste Ziel. (vgl. ebd., S. 26) Für die Meinungsbildung kann dies ein Problem darstellen und zu einer starken Polarisierung in der Gesellschaft führen (vgl. Humborg & Nguyen, 2018, S. 21).

Diese Fokussierung dieser Algorithmen mit ihren Empfehlungssystemen auf die angeklickten Inhalte mit der Personalisierung der Systeme und unserer Tendenz zu homogenen sozialen Freundeskreisen ist es, die einigen Menschen Sorgen bereitet. Denn die Frage, die sich hierbei gestellt werden muss, ist, was mit NutzerInnen in einem sozialen Netzwerk passiert, wenn diese sich erneut immer wieder Nachrichten eines speziellen Typs ansehen? Es kann hier beispielsweise angenommen werden, dass jemand ein Fan von mehreren Musikbands und SängerInnen unterschiedlicher Genres ist, verschiedene Musikmagazine liest und selbst auf der Onlineversion seiner Tageszeitung die Kulturberichte bevorzugt. Es ist hier davon auszugehen, dass bei einer starken Personalisierung dieser NutzerInnen mit der Zeit nur mehr Nachrichten bekommen, die über Kultur & Entertainment handeln. Im Gegensatz zu den klassischen Medien wie Fernsehen, Radio oder der Tageszeitung wäre diese auch anderen Inhalten ausgesetzt, auch wenn sie sich hauptsächlich für den Kulturteil interessieren würden. Im Hinblick auf die politische Meinung ist hier die Sorge groß, dass NutzerInnen, die sich hauptsächlich mit einer extremen politischen Meinung auseinandersetzen, auch nur mehr solche Inhalte gezeigt bekommen, die ihre extreme Meinung unterstützen. Hierbei ist die Meinungsbildung durch den starken Algorithmus im Filter also nicht mehr so frei wie in den klassischen Medien. Dieses Phänomen wird, wie es von Eli Pariser erstmalig beschrieben wurde, „Filterblase“ genannt. (vgl. Deussen, Krafft, & Zweig, 2017, S. 323f.)

Das ist jedoch nicht der einzige Effekt, den die Algorithmen sozialer Medien mit sich bringen: Es ist bekannt, dass wir Menschen uns vor allem mit Menschen wohlfühlen, die uns in verschiedenen Sichtweisen ähnlich sind. Zumeist ist es also der Fall, dass Personen aussucht werden, die uns im Alter, Bildungs- oder

Familienstand ähneln und möglichst auch die gleiche politische Meinung vertreten. Die Empfehlungssysteme der sozialen Medien merken beziehungsweise lernen, mit welchen Personen man sich gerne umgibt und versuchen so ähnliche Bekannte und Freunde vorzuschlagen. Im Hinblick dessen bekommt man also hauptsächlich Nachrichten und Informationen, die Meinungen vertreten, die man schon unterstützt oder die eine Meinung ergänzt. Durch das unbewusste Umgehen mit NutzerInnen ähnlicher Meinung kommt es zu einer einseitigen Verlagerung der Realität. Im Gegensatz zum echten Leben, in dem es sich nicht vermeiden lässt auch mit anderen Leuten, die jemanden nur wenig ähnlich sind, in Kontakt zu treten, wird sich im Internet in sogenannten geschlossenen „Echokammern“ bewegen. Menschen umgeben sich hauptsächlich mit Personen, die dieselbe Meinung vertreten. Dies führt zu einer Verzerrung der Realitätssicht. Durch die ständige Bestätigung der eigenen Meinung entsteht bei NutzerInnen der Eindruck, als würden sie zu der Mehrheit gehören. Hinsichtlich extremer Meinung kann dies zu fragwürdigen Konsequenzen führen. (vgl. ebd., S.324) Diese Phänomene der sozialen Netzwerke beeinflussen Personen dahingehend, da das Meinungsklima anderer sowohl die Redebereitschaft als auch die Meinungsbildung von Individuen beeinflusst. (vgl. Schweiger, 2017, S. 128)

## 2.5 Zusammenfassung

Die sozialen Medien sind heute nicht mehr aus dem Leben vieler Menschen wegzudenken. Das Teilen von Informationen sowie das Kommunizieren mit anderen Personen wurde dadurch wesentlich erleichtert. Die Bevölkerung informiert sich heute nicht mehr nur über die klassischen Medien, sondern auch über die sozialen Medien. Medienunternehmen veröffentlichen heute Informationen nämlich auch in den sozialen Netzwerken und auch NutzerInnen teilen Informationen mit anderen Personen. Daher wird von einer erweiterten Medienöffentlichkeit sowie von einer veränderten persönlichen Öffentlichkeit gesprochen. Die Öffentlichkeiten wurden jedoch nicht nur verändert, sondern auch miteinander vernetzt. Zudem hat sich auch die Kommunikation maßgeblich verändert. Es ist mittlerweile nicht mehr notwendig ständig auf die Nachrichten im Fernsehen oder im Radio zu warten. Die aktuellsten Nachrichten werden heute

hauptsächlich über die sozialen Medien verbreitet. Außerdem können NutzerInnen schnell und einfach auf Informationen reagieren. Es ist heute sehr einfach am Kommunikationsgeschehen teilzunehmen und einzugreifen in dem Personen selber zum Sender werden und sich uneingeschränkt der Welt mitteilen können. Dadurch kommt es jedoch immer wieder zu der Verbreitung von „fake news“ und „hate-speech“. Neben dem Einfluss auf die Öffentlichkeit sowie auf die Kommunikation, beeinflussen soziale Medien auch die Meinungsbildung. Durch sogenannte „Filterblasen“ und „Echokammern“ befinden sich NutzerInnen in einer Umgebung, in der nur Informationen erscheinen, die den Interessen entsprechen und wo nur jene Personen mit ähnlicher Meinung zum Freundeskreis zählen. Dadurch wird der Eindruck erweckt, als gehöre man der Mehrheit an, so dass es zu einer verzerrten Realitätssicht kommt.

### 3 Hate Speech in den Sozialen Medien

In diesem Kapitel geht es um das Phänomen der Hate Speech. Dabei wird zu Beginn der Begriff selbst und daraufhin dessen Auswirkung auf die Gesellschaft erläutert. Abschließend werden wichtige Zahlen und Fakten dazu aufgezählt sowie auf mediale Beispiele bezüglich Hate Speech eingegangen.

#### 3.1 Was ist Hate Speech?

In mediatisierten Wellen der Empörung und Moralpaniken werden Namensgebungen beziehungsweise Bezeichnungen gefunden, die verletzend, verunglimpfend, herabwürdigend, sowie beleidigend sein können. Hate Speech repräsentiert deshalb nicht nur Gewalt, sondern solch eine gewaltvolle Sprache ist Gewalt. (vgl. Turner, 2017, S. 119)

Hate Speech weißt einige Gemeinsamkeiten mit Konzepten wie den Vorurteilen und eben der Gewalt auf. (vgl. May, 2018, S. 400)

Laut Sponholz (2018) handelt es sich bei Hate Speech nämlich um die öffentliche Kommunikation bewusster beziehungsweise intentionaler Botschaften mit diskriminierenden Inhalten. Zentral handelt es sich dabei um eine Art der kommunikativen Herstellung menschlicher Minderwertigkeit. (vgl. Sponholz, 2018, S. 48) Es wird bewusst Antipathie ausgedrückt, die sich auf ein Individuum bezieht, das einer bestimmten Gruppe zugehörig ist oder richtet sich generell gegen eine gewisse Gruppe von Menschen. (vgl. Simpson, 2013, S. 701) Häufig sind dabei MigrantInnen, Frauen, MuslimInnen, JüdInnen oder Homosexuelle betroffen. Diese Gruppen werden als Minderheiten definiert und besitzen meist eine ungünstige politische oder soziale Stellung im gesellschaftlichen Machtgefüge. (vgl. Sponholz, 2018, S. 48)

Vor allem organisierte Gruppierungen versuchen mit der gezielten Verbreitung von Hate Speeches Minderheiten zu unterdrücken und ihnen einen bestimmten sozialen Status zuzuschreiben. (vgl. Whitten, 2019, S. 4)

Hate Speech gilt als die momentan bekannteste Ausprägung verbalisierter Missachtung im Internet. Durch die sozialen Medien, bei deren Inhalten es sich häufig um „User-Generated-Content“ handelt, der durch ein hohes Maß an Interaktionen gekennzeichnet ist, haben sich die Möglichkeiten vermehrt

Verachtung zum Ausdruck zu bringen. Der Hass im Internet wird als vernetzt, interaktiv und multimedial beschrieben. (vgl. Ganz, 2019, S. 2)

Obwohl schon während der Zeit des Nationalsozialismus solche Hate Speeches beobachtet werden konnten, ist der Ausdruck doch relativ neu. Erst in den 80er Jahren konnte dieser Begriff in den Archiven von Google gefunden werden. Seitdem hat sich dieser rasant ausgeweitet. Diese schnelle Entwicklung ist der Grund dafür, warum es weder in der Wissenschaft noch im Gesetz eine einheitliche Definition von Hate Speech gibt. (vgl. Sponholz, 2018, S. 48f.) Es ist aber nicht nur die Definition uneinheitlich, sondern auch der Ausdruck selbst. Der Begriff Hate Speech ist nicht notwendigerweise von Hass getrieben oder auf sprachliche Äußerungen begrenzt. Daher kann gesagt werden, dass „Hassrede“ als deutsche Übersetzung irreleitend beziehungsweise falsch ist. Vielmehr kommt das deutsche Wort „Hetze“ dem Phänomen der Hate Speech näher. Dieses Wort hat aus juristischer Perspektive jedoch den Beigeschmack an seinen gesellschaftlichen Folgen gemessen zu werden. Die Folgen von Hate Speeches können jedoch sehr vielfältig sein, besonders für die betroffenen Personen selbst. (vgl. ebd., S. 50)

Die Bezeichnung Hate Speech zeigt eine definitorische Unschärfe auf. Es wird oft mit dem Wort Hass in Verbindung gebracht und als emotionales Phänomen behandelt. Hass ist etwas, das in dem Inneren eines Individuums entsteht. Folgt man dieser Definition, so wäre die Kommunikation menschenverachtender Ideologien wie damals während des Nationalsozialismus eine individuelle Angelegenheit. Es ist also festzuhalten, dass sich das Wort „Hate“ in diesem Fall nicht auf die Emotion Hass bezieht, sondern auf Feindlichkeit gegenüber Menschen einer bestimmten Kategorie. (vgl. ebd., S. 51ff.)

Auch das Wort „Rede“ wird häufig falsch verstanden und zu meist mit Wörtern assoziiert. Jedoch können Hate Speeches auch völlig auf Worte verzichten. Auch deswegen ist die deutsche Übersetzung irreführend und falsch. Beispielsweise wird auch beim Aufstellen brennender Kreuze vor den Häusern afroamerikanischer Familien in den USA von einer Hate Speech gesprochen. In den sozialen Medien sind es ebenfalls häufig nur Bilder, die als Wörter verwendet werden und so feindliche Reaktionen oder Hasspostings auslösen. Bei Hate Speeches geht es um Botschaften, die nicht gezwungenermaßen durch das

Gesprochene übermittelt werden. Die Sprache ist nur ein möglicher Code. (vgl. ebd., S. 57)

Wichtig zu erwähnen ist, dass Hate Speeches nicht mit Beschimpfungen oder Beleidigungen gleichzusetzen sind. Im Gegensatz dazu sind die Merkmale von Hate Speeches die öffentliche Kommunikation, gruppenbezogene Diskriminierung oder Wertung und die Intentionalität. (vgl. ebd., S. 81.ff) Die folgende Tabelle stellt die wesentlichen Unterschiede zu anderen Formen der Abwertungen dar:

Merkmale	Diskriminierende Sprache	Hate-Speech	Beschimpfung	Beleidigung	Fauxpas
Öffentlichkeit	möglich	ja	möglich	möglich	ja
„Gruppen“-bezug/Kategorisierung	ja	ja	möglich	möglich	möglich
Evaluation/Wertung	ja	ja	ja	ja	ja
Intentionalität	möglich	ja	ja	möglich	nein
Begründungs-rationalität	möglich	möglich	nein	nein	nein
Lexikalisierung	möglich	möglich	ja	möglich	möglich
Performativität/Wirkung	möglich	möglich	möglich	ja	ja

*Tabelle 2: Formen symbolischer Abwertungen – Eigene Darstellung nach Sponholz 2018, S. 83*

Ebenfalls zu unterscheiden ist der Begriff der Hate Speech mit den Hasspostings. Während sich Hate Speeches auf bestimmte Gruppen beziehen, beschränken sich Hasspostings nicht auf kategoriebezogene Aussagen. Bei den sogenannten Hasspostings ist lediglich Cyber Hate gemeint. (vgl. Wettstein, 2016) Es gibt jedoch auch Hasspostings, die sich auf eine bestimmte Kategorien von Menschen beziehen. Dabei handelt es sich in der Regel um eine spezielle Form von Hate Speech, nämlich der Hateful Speech. (vgl. Sponholz, 2018, S. 85)

Des Weiteren gibt es die Unterscheidung zwischen Hate Speech und Political Incivility. Hierbei geht es um Beschimpfungen von PolitikerInnen und/oder zwischen NutzerInnen (vgl. Conway, Danielson, & Stryker, 2016, S. 540). Dazu zählen beispielsweise abwertende Spitznamen, die Nutzung eines vulgären Wortschatzes und die zielgerichtete Behauptung von Unwahrheiten (vgl. Sponholz, 2018, S. 87). Obwohl beide dieser Phänomene eine sehr hohe

Relevanz haben sind diese normalerweise nur in Ausnahmefällen mit Hate Speech gleichzusetzen. Bei der Political Incivility geht es um die gegenseitige Positionierung von zwei oder mehreren AkteurInnen. Im Vergleich dazu werden bei der Hate Speech zwei Kategorien von Menschen als Gegensätze definiert. (vgl. ebd.)

Obwohl Sponholz (2018) den Begriff Hate Speech hier klar definiert und von anderen Begriffen abgrenzt, fällt dieser im rechtlichen und sozialen Verständnis sehr unterschiedlich aus. (vgl. Der Standard, 2018) Aus diesem Grund wird der Begriff Hate Speech hier als Dachbegriff für Hasspostings bzw. Hateful Speech sowie für Political Incivility verwendet.

Wenn über Hate Speech gesprochen wird wird zwischen zwei verschiedene Typen von unangenehmen InternetuserInnen unterschieden: Zum einen gibt es den sogenannten „Troll“ und zum anderen den „Glaubenskrieger“.

Ein Troll ist jemand, der vorgibt ein Teil einer professionellen Gruppe zu sein, in Wahrheit jedoch nur stört und somit Konflikte hervorrufen möchte. (vgl. Hardaker, 2010, S. 216f.)

Bei dem Begriff des „Trolles“ wird im Deutschen sehr leichtfertig mit diesem Begriff umgegangen. Oft wird schon jemand als Troll bezeichnet, wenn dieser jemanden widerspricht oder beleidigt. Dies ist jedoch ein Missverständnis, denn der Troll geht meist viel berechnender als andere taktlose UserInnen im Internet vor. Der Begriff des Trolls kommt von dem englischen Wort „Trolling“, wenn ein Köder ins Wasser geworfen wird und mit dem Boot davonfährt. Dies ist auch bei dem Internettroll der Fall. Auch dieser wirft einen Köder aus, beispielsweise eine verletzende Wortmeldung. Hier hofft er, dass wir als normale UserInnen anbeißen, denn dies amüsiert den Troll, da dies für ihn ein Beweis ist, dass dieser intelligenter ist als seine Beute. Das Ziel des Trolles ist es also grundsätzlich andere NutzerInnen zur Weißglut zu bringen. Dies erzielen sie damit, indem sie vermeintlich naive Fragen stellen, vom Thema abschweifen, mit überzogener Kritik reagieren oder bewusst eine Gegenposition einnehmen. Des Weiteren haben sie es als Ziel mit Worten, Bildern oder Videos andere NutzerInnen zu schocken und zu provozieren. (vgl. Brodnig., S. 39ff.)

Im Gegensatz dazu gibt es die „Glaubenskrieger“ die online meist für viel Hass verantwortlich sind. Diese möchten zumeist polarisieren indem diese es nicht als notwendig betrachten Andersdenkende mit Respekt zu behandeln. Sie beanspruchen meist die Wahrheit für sich und sehen es als ihre Pflicht an ihre „Wahrheit“ weiter zu verbreiten. Grundsätzlich haben diese wenig Empathie für Betroffene und empfinden sich meist selbst als Opfer. Außerdem stehen sie der Demokratie meist kritisch und ablehnend gegenüber. Glaubenskrieger sind politisch meistens nicht links oder rechts orientiert, obwohl diese Bewegungen für populistische Parteien eine interessant und leicht erreichbare Wählergruppe darstellen. Diese misstrauen meist etablierten Medien. Vor allem die „Lügenpresse“-Rhetorik, die rechtspopulistische Parteien in Österreich und Deutschland prägen, werden von Glaubenskrieger oft als Methode verwendet. Während Trolle sich schnell einmal zurückziehen, wenn die Provokation nicht fruchtet, machen die Glaubenskrieger meist weiter, da diese sich andauernd im recht fühlen und auch kaum Konsequenzen fürchten. Aus diesem Grund ist die Verteidigung hier sehr schwer, da diese sich selten zurückziehen, wenn ihre Attacken ignoriert werden. Ein weiterer Unterschied zum Troll ist auch, dass diese meist Spaß an provozierenden Diskussionen haben, während Glaubenskrieger sich bedroht fühlen oder von einer Angst beziehungsweise einer bestimmten Ideologie getrieben werden. (vgl. ebd., S. 54ff.)

	Trolle	Glaubenskrieger
<b>Empathie</b>	Niedrig	Niedrig
<b>Motivation</b>	Schadenfreude (die LULZ = „ich tat es um mich zu amüsieren“)	Gefühlte Bedrohung
<b>Sieht sich selber als</b>	intellektuell überlegen	heldenhaft
<b>Ideologie?</b>	Nichts ist heilig	Unbeirrbare Überzeugung
<b>Ziel Vorrangige Methode</b>	Menschen manipulieren Provokation	Menschen „wachrütteln“ Panikmache
<b>Angriffspunkt</b>	Frustriert, wenn sich Gegenüber nicht provozieren lässt	Genau betrachtet nicht zur differenzierten Diskussion bereit.

Tabelle 3: Übersicht Unterschied zwischen Troll und Glaubenskrieger - Eigene Darstellung nach Brodnig 2016, S. 75

### 3.2 Die Unterscheidung zwischen online/offline Hate Speech

Bevor im nächsten Kapitel auf die Auswirkungen von Hate Speeches eingegangen wird, wird hier noch kurz auf die Spezifika der online Hate Speeches eingegangen und wie sich diese von den offline Hate Speeches unterscheiden. Es wird hier erläutert, warum online Hate Speeches gegenüber den offline Hate Speeches so speziell sind. Dies wird auch deshalb aufgezeigt um besser zu verstehen, warum die Bekämpfung beziehungsweise Regulierung von einem sogenannten „Cyber Hate“ so herausfordernd ist. (vgl. Brown, 2018, S. 297)

John Suler von der amerikanischen Rider University erkannte bereits vor mehr als zehn Jahren das Gefühl der Unsichtbarkeit im Internet. Damals entwickelte er die Theorie der „Online Disinhibition Effects“, also der „Online-Enthemmungs-Effekt“. Manchmal kann dies etwas Gutes sein, zum Beispiel wenn ein homosexueller Bursche oder Mädchen im Internet zum ersten Mal über ihre Gefühle sprechen. Meistens sehen wir aber die Schattenseite davon: die „toxische Enthemmung“. Hierbei toben sich die UserInnen online aus, ätzen herum und vergiften somit das Diskussionsklima. Hier wird von dem Ausleben niederträchtiger Bedürfnisse und Wünsche ohne irgendein persönliches Wachstum gesprochen. (vgl. Brodnig, 2016, S. S. 13f.) Ein Faktor, der diese Enthemmung im Internet fördert, ist die Anonymität. Dadurch ist die Möglichkeit gegeben zu sagen was man will was man sich bei einer face-to-face Kommunikation möglicherweise nicht trauen würde. Dadurch, dass andere den eigenen Namen nicht kennen, fühlt man sich nicht so verwundbar. Außerdem würden diese bei einer face-to-face Kommunikation das Risiko eingehen von Individuen angegriffen zu werden. Da der Hate Speaker online nicht körperlich anwesend ist müssen sich diese keine Angst vor körperlichen Angriffen haben. (vgl. Brown, 2018, S. 298f.)

Es wurde bereits erwiesen, dass diese Anonymität, die im Internet gegeben ist, Menschen dazu ermutigt fieser, abscheulicher und noch gehässiger zu sein als im wahren Leben. Außerdem erweckt es bei Menschen teilweise den Eindruck, als würden die normalen Verhaltensregeln im Internet nicht gelten. Auf der anderen Seite fällt es hier auch den Opfern von Hate Speeches leichter, sich gegenüber diesen zu äußern und zu wehren. (vgl. ebd.)

Des Weiteren herrscht eine Unsichtbarkeit, die jedoch nicht mit der Anonymität gleichgesetzt werden darf. Bei der Unsichtbarkeit, wird nicht der reale Name verborgen, wie bei der Anonymität, sondern es fallen hier die nonverbalen Signale weg. Dies erklärt unter anderem, warum NutzerInnen auf Facebook unter ihrem realen Namen die schlimmsten Aussagen tätigen. (vgl. Brodnig, 2016, S. 13)

Ein weiteres charakteristisches Merkmal von online Hate Speeches ist, dass es hier zwischen den SprecherInnen und dem Publikum eine gewisse physische Distanz gibt. Das heißt, dass die SprecherInnen für das Publikum nicht sichtbar sind. Für richtige „hardcore“ Hate Speaker würde es jedoch auch nichts ändern, wenn diese Distanz nicht gegeben wäre. Viele genießen es nämlich während ihrer Rede in die leidenden Gesichter der angesprochenen Personen zu blicken. Durch diese Distanz, die das Internet mit sich bringt, schützen diese sich lediglich vor körperlichen Angriffen. (vgl. Brown, 2018, S. 300f.)

Zudem ist es möglich, dass Hate Speeches im Internet oftmals mehr Zuspruch bekommen als im realen Leben. Dadurch, dass sich die Menschen hinter den Bildschirmen verstecken können, scheuen sie sich nicht ihre Meinung zu äußern und Teil umstrittener Communities zu werden. Jeder umgibt sich gerne mit Menschen, die dieselben Meinungen haben. Durch die Anonymität im Internet können Menschen sich einer Community anschließen, die eine diskriminierende Einstellung gegenüber einer bestimmten Gruppe teilen und häufig Hate Speeches im Internet verbreiten. Solche Communities wollen sich vergrößern und noch mehr Menschen erreichen. Durch das Internet ist dies leicht und ohne viel Aufwand möglich. (vgl. ebd., S. 301f.)

Eine weitere Eigenschaft ist, dass es das Internet ermöglicht binnen weniger Minuten Informationen zu veröffentlichen und damit schnell ein großes Publikum zu erreichen. Während man bei einer face-to-face Kommunikation nur eine Person oder eine Gruppe erreicht, können Hate Speeches im Internet gezielt und mehrere Gruppen gleichzeitig verbreitet werden. (vgl. ebd., S. 304)

Wer einen hasserfüllten Kommentar im Internet verfasst, bekommt oft kein unmittelbares Feedback. Das heißt, dass man sich nicht sofort damit beschäftigen muss, was die eigenen Worte anrichten. Hier wird von der sogenannten Asynchronität gesprochen. ExpertInnen sprechen bei diesem Phänomen auch von einer „emotionalen Fahrerflucht“. (vgl. Brodnig, 2016, S. 13)

Durch all diese Eigenschaften entgleisen Debatten im Internet häufig sehr schnell. Auch Menschen, die normalerweise umgänglich sind, diskutieren online energisch und zeigen eine ungewohnte schroffe Wortwahl. Schimpfworte und andere Beleidigungen scheinen online schneller über die Lippen beziehungsweise über die Tastatur zu rutschen. Es wird heute zumeist schriftlich über das Smartphone oder den Computer diskutiert. Der Augenkontakt, die Mimik und die Gestik sowie die Stimme der GesprächspartnerInnen, die doch so entscheidend sind, fehlen hier jedoch. Dadurch gehen wichtige Informationen verloren, denn nonverbale Signale fördern die Empathie. Wenn man ein persönliches Gespräch mit einer Person führt, erkennt man normalerweise am Gesichtsausdruck des Gegenübers, ob man in der Tonalität zu extrem ist. Genau diese nonverbalen Signale fehlen in der Online-Kommunikation und so auch die Empathie. (vgl. ebd., S. 13.ff.)

### **3.3 Die Auswirkungen von Hate Speech**

Öffentliche Hate Speeches haben den Zweck bestimmte Zielgruppen direkt zu attackieren. Dies führt dazu, dass Opfer ihre Einstellungen ihnen selber, bestimmten Gruppen und der Gesellschaft gegenüber verändern. (vgl. Fumagalli, 2019, S. 7) Seglow (2016) behauptet zum Beispiel, dass öffentliche Hate Speeches die Selbstachtung des Opfers negativ beeinflussen. Laut Waldron (2012) verändern diese auch den Geisteszustand des Opfers bezüglich ihrer Würde als eine gleichberechtigtes Mitglied einer politischen Gemeinschaft. Des Weiteren führen öffentliche Hate Speeches zu einem psychologischen Trauma, zu physischen Symptomen sowie zu emotionalen Schmerz. Es kann dazu kommen, dass das Opfer beispielsweise unter ständigen Bauchschmerzen, einer rasenden Impulsrate, Schwierigkeiten in der Atmung, Albträume oder einer Psychose leidet. Teilweise führen solche Ereignisse auch bis zum Selbstmord des Opfers. (vgl. Matsuda, 2018, S. 24)

Öffentliche Hate Speeches greifen an, entwürdigen und schaden Zielgruppen indem diese beginnen an ihrer mentalen Einstellung zu zweifeln. (vgl. Fumagalli, 2019, S. 8) Laut Delgado (2018) ist ein direkter und negativer Effekt auch damit verbunden, ob ein Opfer schon in einem Umfeld leben, in dem Diskriminierungen sehr verbreite und üblich sind.

Jede Person in einer Gesellschaft hat eine Würde, die von anderen zu respektieren ist. Es wird von einem gesetzlichen Status gesprochen, da alle in einer liberalen Gesellschaft die gleichen nationalen Rechte genießen. Die Menschenwürde ist auch ein ziviler Status, da alle BürgerInnen vollberechtigte Mitglieder einer Gesellschaft sind und sollten auch als solches respektiert werden. Es ist also ein soziales Ziel diese Würde eines Einzelnen in einer Gesellschaft zu respektieren und aufrechtzuerhalten. (vgl. Seglow, 2016, S. 1106) Stellt man sich nun eine Gesellschaft vor die gesättigt ist von Hate Speeches, so würden Minderheiten als bestialisch und unmenschlich oder als kriminelle und perverse Menschen dargestellt werden. Diese würden in unserer Gesellschaft nicht mehr respektiert und nicht mit der Mehrzahl gleichgestellt werden (vgl. Waldron, 2012, S. 66) Geht eine Gesellschaft gegen solche Aktionen nicht vor, so können Minderheiten nicht mehr auf die mehrheitliche Hingabe der Gleichberechtigung vertrauen. Dies würde zu einer nicht-geordneten Gesellschaft führen. (vgl. Rawls, 1999, S. 397ff.)

Eine geordnete Gesellschaft in diesem Sinne ist notwendig, so dass Menschen die Wege gehen können, die sie wollen. Dadurch fühlen sie sich sicher und können soziale Interaktionen ohne jegliche Risiken entgegensehen. (vgl. Waldron, 2012, S. 83f.) Laut Waldron ist es also notwendig innerhalb einer Gesellschaft eine Versicherung als öffentliches Gut zu haben, das dafür sorgt, dass die Menschenwürde eines Einzelnen respektiert wird. Nur so ist ein geordnetes und friedliches Zusammenleben möglich. (vgl. ebd., S. 94f.)

Im Gegensatz dazu haben Hate-Speeches auch indirekte Auswirkungen. Durch den Inhalt von diesen Reden wird versucht das Publikum von bestimmten Ansichten zu überzeugen beziehungsweise eine neue Denkweise anzuregen. Beispielsweise wird versucht den Menschen die Idee des Rassenhasses und der Diskriminierung glaubhaft zu machen. Es wird versucht das Publikum so zu überzeugen, dass so ein schädliches Verhalten und Gewalt gegenüber bestimmten Zielgruppen normal sei. (vgl. Fumagalli, 2019, S. 10)

Laut Langton (2012) wissen Hate Speaker häufig sehr gut, wie diese ihr Publikum von ihren Ansichten überzeugen können, so dass diese ihre mentale Einstellung gegenüber einer bestimmte Gruppe ändern.

Vor allem in den sozialen Medien stellt dies ein Problem dar. Nicht nur, weil dadurch sehr schnell sehr viele Menschen auf einmal erreicht werden, sondern auch durch die, wie schon genannten, „Echokammern“ und „Filterblasen“. Wird man nun durch solche Hate-Speeches indirekt beeinflusst, indem sich die Ansichten gegenüber Personengruppen verändern, so verändern sich auch die Interessen. In den sozialen Netzwerken gerät man so in ein anderes Umfeld, dass sich hauptsächlich mit diskriminierenden Inhalten befasst. Es entsteht also eine neue Realität für die Betroffenen. (vgl. Deussen et al., 2017, S. 323f.) Können nun immer mehr Leute von solchen fragwürdigen Denkweisen überzeugen, so entsteht, wie oben erwähnt, eine nicht-ordentliche Gesellschaft. (vgl. Rawls, 1999, S. 397ff.)

### **3.4 Zahlen und Fakten zu Hate Speech**

Im Jahr 2015 hat das No-Hate-Speech-Movement eine Online-Erhebung mit 6.601 jungen Leuten durchgeführt. Dabei kam heraus, dass 83% davon schon einmal Hass im Netz erlebt haben. (vgl. BMSGK, 2019) Laut der Beratungsstelle #GegenHassimNetz, die von dem Verein ZARA betrieben wird, wurden allein im letzten halben Jahr nur in Österreich mehr als 500 Fälle von Online-Hass und Hetze und Cyber-Mobbing dokumentiert. Davon verstößen mehr als 40% Prozent gegen den Verhetzungsparagraphen oder das NS-Verbottgesetz, dass im Jahr 2016 eingeführt wurde. (vgl. Kurier, 2018)

Im Rahmen einer Online-Befragung des Instituts für Strategieanalysen, bei der 1.204 Personen in Österreich im Alter von 14 bis 70 Jahren befragt wurde, konnte herausgefunden werden, dass 18% sehr oft Hate Speeches im Internet unterkommen. 27% der Befragten haben dabei mit „ja, manchmal“ und 22% mit „ja, aber nur selten“ geantwortet. (vgl. ISA, 2017, S. 54)

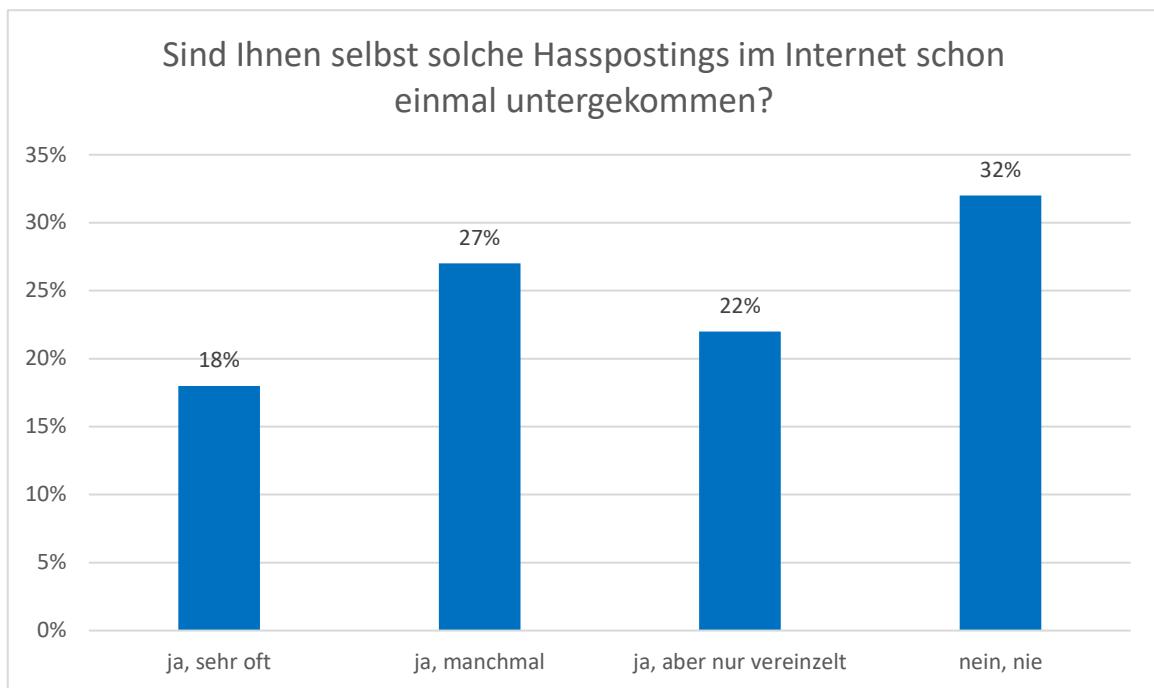


Abbildung 1: Erfahrungen mit „Hass-Postings“ - Eigene Darstellung nach ISA 2017, S. 54

Das heißt, dass 67% schon Hasspostings im Internet untergekommen sind. Lediglich 32% sind solchen im Internet noch nie begegnet. Vor allem die unter 19-Jährigen und 19- bis 29-Jährigen begegnen häufig solchen Postings. (vgl. ebd.) Bei der Frage wie auf diese Hasspostings reagiert wird, antworteten die meisten der Befragten damit, dass sie diese ignorieren. 31% melden diese den ModeratorInnen beziehungsweise nutzen eine Meldefunktion und 29% verlassen die Seite oder das Forum. (vgl. ebd., S. 55)

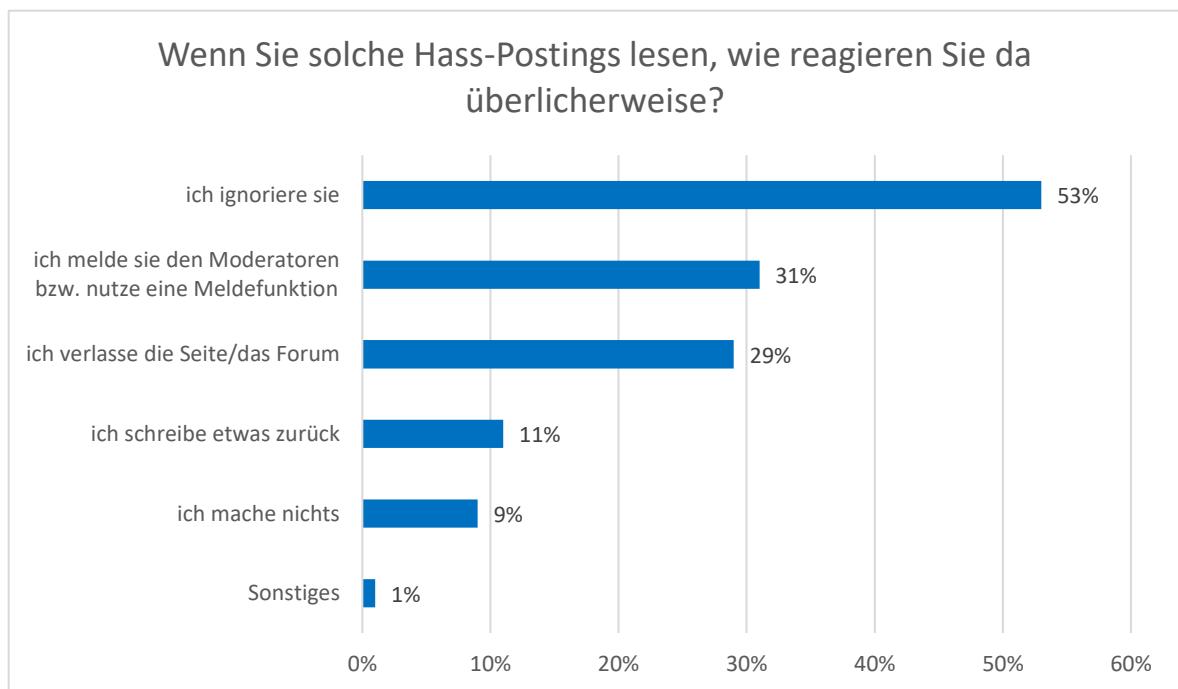


Abbildung 2: Reaktionen auf "Hass-Postings" – Eigene Darstellung nach ISA 2017, S. 55

Ähnlich sieht es im Nachbarland Deutschland aus. Hier kam bei einer Forsa-Studie heraus, dass 78% der Befragten schon einmal eine Hassrede beziehungsweise Hasskommentare im Internet gesehen haben (z.B. auf Webseiten, in Blogs, in Internetforen oder auf sozialen Netzwerken). Nur jeder fünfte Befragte, also 22%, sind bisher noch keiner Hassrede beziehungsweise Hasskommentaren im Internet untergekommen. Vor allem Befragte im Alter zwischen 14 und 24 Jahren sehen häufiger Hassreden beziehungsweise Hasskommentare. (vgl. Forsa., 2018, S. 1ff.)

Bei der Frage, wie die Befragten sich bei der Wahrnehmung von Hassreden/Hasskommentaren im Internet verhalten, sagten 37%, dass diese sich näher damit befassen. 26% antworteten damit, dass diese solche hasserfüllten Reden im Internet melden und 25% gaben an auf solche Reden beziehungsweise Kommentare zu antworten und zu kritisieren. 42% der Befragten hingegen haben angegeben, sich weder näher mit diesen hasserfüllten Reden und Kommentaren zu befassen oder VerfasserInnen zu melden, noch auf diese geantwortet zu haben. (vgl. ebd.)

Mit der Frage, welche Gruppen und Personen besonders von Hate Speech betroffen sind, hat sich eine hessische Studie befasst. Laut dieser deutschen Untersuchung beobachteten die befragten Personen am häufigsten aggressive oder abwertende Aussagen gegenüber Menschen mit Migrationshintergrund, geflüchteten Menschen und MuslimInnen. Außerdem haben sie auch häufig hasserfüllte Aussagen gelesen, die sich auf die Bundesrepublik Deutschland, Amtierende PolitikerInnen oder politisch Andersdenkende beziehen. (vgl. Eckes et al., 2018, S. 20)

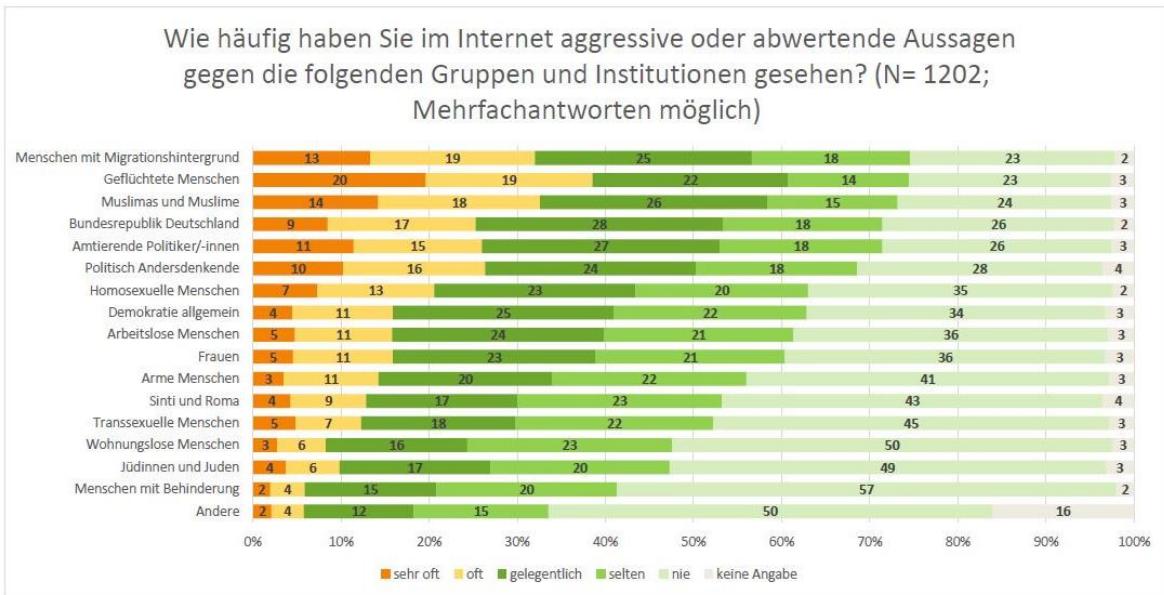


Abbildung 3: Gegen welche Gruppen und Institutionen richtet sich der Hass (Eckes et al. 2018, S. 20)

Eine Analyse von der Beratungsstelle #GegenHassimNetz zeigt ebenfalls, dass sich auch in Österreich die meisten Hasspostings auf Moslems und Flüchtlinge beziehen. Viele Hasspostings richten sich jedoch auch gegen Frauen. Anhand dieser Analyse wurde ebenfalls ersichtlich, dass auf Facebook die meisten Hasspostings gefunden und gemeldet wurden. Auch auf Twitter und YouTube fand man hetzerische Online-Artikel und Kommentare sowie verletzende Inhalte. (vgl. Kurier, 2018)

Laut des Rassismus Reports von ZARA kommen rund 60% aller Meldungen, die Rassismus im Internet betreffen, von Facebook-UserInnen. Daraus kann man schließen, dass das Bewusstsein, rassistischen Hass an die Beratungsstelle zu melden, auf Facebook sehr hoch ist. (vgl. ZARA, 2018, S. 13)

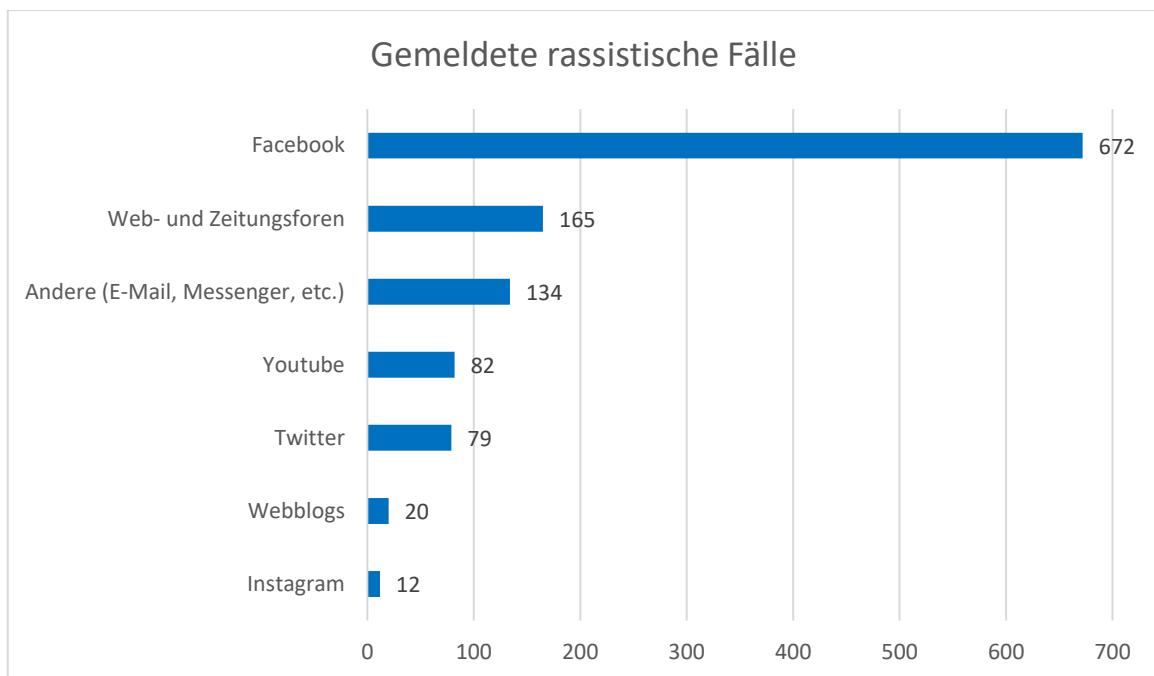


Abbildung 4: *Gemeldete rassistische Fälle – Eigene Darstellung nach ZARA 2018, S. 13*

Auf Facebook lässt sich vor allem beobachten, dass Hass ansteckend ist. Meist sind dort die Ausgangspostings noch vergleichsweise harmlos. In den Diskussionen darunter wird es dann jedoch sehr schlimm. In einigen Debatten hagelt es nur so von rassistischen Kommentaren. (vgl. Brodnig, 2016, S. 7ff.) Im Fall von rassistischem Hass sind fast zwei von drei Fällen in Form eines Kommentars verbreitet. (vgl. ZARA, 2018, S. 13)

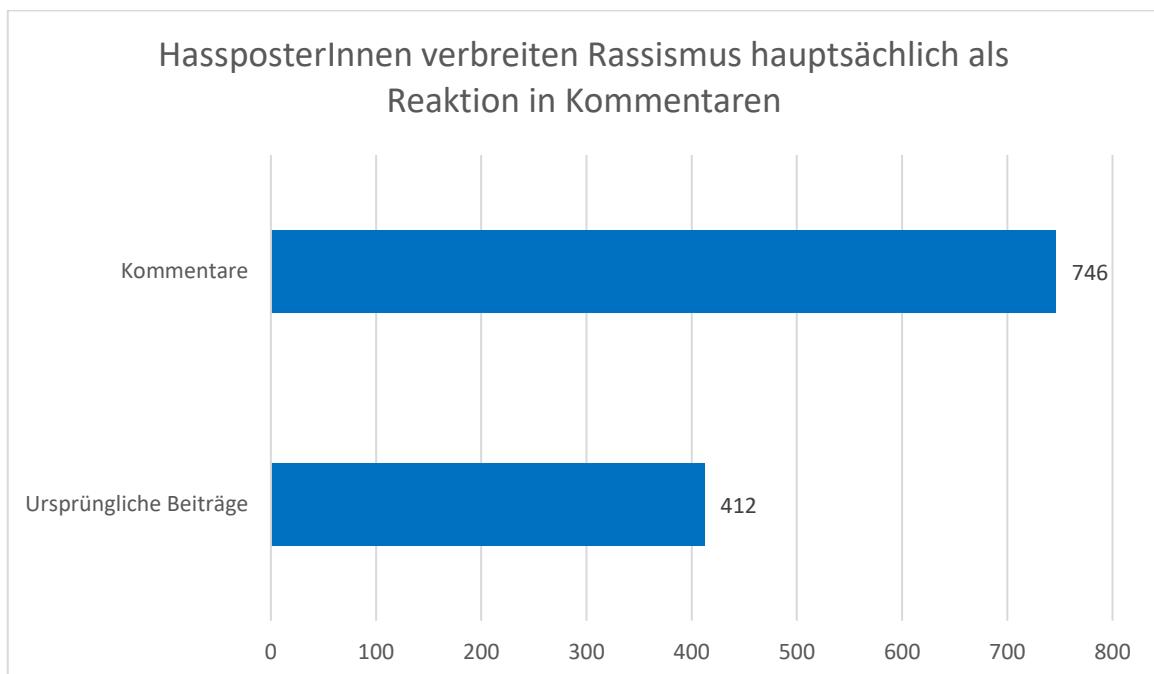


Abbildung 5: *HassposterInnen verbreiten Rassismus hauptsächlich als Reaktion in Kommentaren – Eigene Darstellung nach ZARA 2018, S. 13*

## 3.5 Mediale Beispiele zu Hate Speech

### 3.5.1 FPÖ-Landesrat wegen Hassposting angeklagt

Bei diesem Fall handelte es sich um einen 51-jährigen Niederösterreicher der wegen eines Hasspostings auf der Seite eines oberösterreichischen FPÖ-Landesrates angeklagt wurde. Dabei meinte er, dass Flüchtlinge „so fein verteilt werden sollten, dass diese nur mehr auf molekularer Ebene nachweisbar sind“. Obwohl der Unternehmer sich vor Gericht damit rechtfertigte, dass er bei seinem Posting nicht zum Umbringen von Flüchtlingen aufgerufen habe, sondern vielmehr dafür eingetreten ist, diese gerecht zu verteilen, wurden Flüchtlinge in diesem Posting sinngemäß als „wilde Tiere“ und als „faul“ bezeichnet. Für die Staatsanwaltschaft und auch die Richterin handelte es sich hier um ein „typisches Hassposting“. Da der Angeklagte immer einsichtiger wurde und wegen seiner bisherigen Unbescholtenheit war eine Diversion möglich. Nach der Bezahlung eines Kostenbetrags von 150€ wurde das Verfahren ohne Verurteilung eingestellt. (Die Presse, 2019)

### 3.5.2 FPÖ-Politiker wegen Hassposting aus Partei geworfen

Ebenfalls diskriminierend war das Hasspostings eines anderen FPÖ-Politikers, der gegen eine Abgeordnete der Grünen hetzte. Dieser hat damals ein Hassposting gegen die Grünen Abgeordnete Alev Korun verfasst. Dabei habe der Kommunalpolitiker auf seiner Facebook-Seite ein Foto von Korun neben einem Bild einer mutmaßlich gesteinigten Frau gepostet und außerdem einer Userin zugestimmt, dass Korun „verbuddelt“ werden sollte. Außerdem beschimpfte er Korun als „wahnsinnige Dreckschleuder aus der Türkei“. Daraufhin brachte die Grünen Abgeordnete eine Privatklage wegen Beschimpfung, übler Nachrede, Kreditschädigung und Ehrenbeleidigung ein. Von einem Grazer Richter wurde daraufhin die unverzügliche Löschung dieses Postings angeordnet. Neben dieser Klage wurde das FPÖ-Mitglied von dem steirischen FPÖ-Landesparteiobmann von der Partei ausgeschlossen. (vgl. Die Presse, 2015)

### **3.5.3 Hassposting gegen Neujahrsbaby**

Neujahr letzten Jahres hatten zahlreiche Medien ein Foto des Neujahrsbaby und deren Eltern veröffentlicht. Die Mutter des Babys trug auf dem Foto ein Kopftuch, das die Familie als muslimischen Glaubens auswies. In den sozialen Netzwerken führte dies zu vielen rassistischen Kommentaren. Einige HassposterInnen konnten von der Strafverfolgungsbehörde ausgeforscht und zur Anklage gebracht werden. Einer dieser Hassposter war ein 65-jähriger Pensionist, der sich auf Facebook zu dem Foto folgendermaßen äußerte: „Für jedes österreichische Baby werden sechs muslimische Jihadisten geboren.“ Dass dies eine Herabsetzung von MuslimInnen darstellte, wollte der Angeklagte nicht einsehen. Obwohl sich dieser zu Beginn noch nicht schuldig bekannte, gab er schlussendlich zu, „eine Dummheit“ begangen zu haben. (vgl. Der Standard, 2019a)

### **3.5.4 Eva-Glawischnig gegen Facebook**

Am bekanntesten ist jedoch der Fall der Grünen beziehungsweise der ehemaligen Grünen-Chefin Eva Glawischnig, die Facebook verklagte, da das soziale Netzwerk Hasspostings nicht löschte. Der Prozess begann im Jahr 2016 als die Grünen juristisch gegen Facebook vorgingen. Dabei ging es nicht nur um die Löschung von Hasspostings, die auf Frau Glawischnig bezogen waren, sondern Ziel war es auch Betroffenen das Vorgehen gegen Hate Speeches zu erleichtern. Vorerst konzentrierte man sich jedoch darauf, dass Hasspostings, die sich auf die damalige Grünen-Chefin Eva Glawischnig bezogen, gelöscht wurden. Bei diesen Postings handelte es sich um Aufrufe wie: „Eva Glawischnig gehöre in die Gaskammer“. Auch Gewaltaufrufe gegen GrünwählerInnen wurden, trotz Meldung der Grünen, nicht gelöscht. Zwar wurden diese Postings in Österreich gesperrt, vom Ausland aus blieben diese jedoch noch abrufbar. (vgl. Schmid, 2016)

Mit dieser Klage konnte eine einstweilige Verfügung gegen Facebook erreicht werden. Facebook musste somit dafür sorgen, dass dieser hasserfüllte Beitrag über Glawischnig weltweit gesperrt wird. Trotz dieses kleinen Erfolges gab es weiterhin mehrere Knackpunkte. Es stellte sich zum einen die Frage, ob Facebook verpflichtet ist auch wortgleiche und sinngleiche beziehungsweise umformulierte Beiträge zu sperren. Außerdem ist fraglich, ob das soziale Netzwerke über solche Hasspostings in Kenntnis gesetzt werden muss oder selbst danach suchen muss.

Deshalb landete das Verfahren vor dem Obersten Gerichtshof, der sich daraufhin an den Europäischen Gerichtshof wandte. (vgl. ebd.)

Vorerst entschied jedoch das Rekursgericht, dass Facebook wortgleiche Beiträge jener Postings, über die die Grünen Facebook informiert hatten, entfernen muss. Die Untersagung sinngleicher Inhalte wurde jedoch abgelehnt. Das Gericht deutete dennoch auf den „Einsatz technischer Hilfsmittel wie eines automationsunterstützten Filtersystems“ hin. 2018 stellte der Oberste Gerichtshof einen Vorlagebeschluss fest, dass Glawischnig „zu Recht“ eine Unterlassung der Verbreitung des Hasspostings gefordert hatte. Die Verpflichtung zu einer Filterung wurde jedoch nicht ausgesprochen. Dies ist deshalb der Fall, da die EU-Richtlinie besagt, dass Mitgliedsstaaten Anbietern von Diensten „keine allgemeinen“ Verpflichtungen auferlegen dürfen, übermittelte oder gespeicherte Informationen zu überwachen oder aktiv nach Rechtswidrigkeiten zu suchen. Zugleich könnten Mitgliedsstaaten jedoch von Dienstanbietern verlangen, eine Sorgfaltspflicht anzuwenden, die bestimmte Arten rechtswidriger Tätigkeiten aufdecken sowie verhindern soll. Um diesen Widerspruch nun zu klären, wurde der Europäische Gerichtshof herangezogen. (vgl. ebd.)

### **3.6 Zusammenfassung**

Hate Speech sind Botschaften mit diskriminierenden Inhalten die öffentlich kommuniziert werden. Hauptsächlich handelt es sich dabei um die kommunikative Herstellung menschlicher Minderwertigkeit. Während vor der Zeit des Internets diese noch persönlich übermittelt wurden, verstecken sich heute die meisten Hate Speaker hinter den Bildschirmen. Durch die Anonymität und die physische Distanz, die das Internet und somit auch die sozialen Netzwerke bieten, trauen sich mehr Leute sich auf Facebook, Twitter & Co. Hate Speeches zu posten. Zumeist handelt es sich bei den Opfern um bestimmte Personengruppen wie Frauen, Homosexuelle, MigrantInnen, MuslimInnen oder JüdInnen. Durch diesen öffentlichen Angriff fühlen sich Opfer in ihrer Würde verletzt und geben ihnen ein Minderwertigkeitsgefühl. Viele erleiden dabei ein psychisches Trauma und emotionalen Schmerz. Es führt aber auch häufig zu physischen Schmerzen wie Bauchschmerzen oder Atemschwierigkeiten. Manchmal führen solche Hate

Speeches die Opfer auch bis zum Suizid. Hate Speeches haben jedoch auch eine indirekte Auswirkung auf unsere Gesellschaft. Durch den Versuch der Verbreitung von Meinungen solcher Hate Speaker kann es zu der Entstehung einer nicht-ordentlichen Gesellschaft kommen. Das heißt, dass nur mehr die Meinung von den Mehrheiten zählen, sodass sich Minderheiten nicht mehr auf diese verlassen können.

Mehrere Studien haben mittlerweile aufgezeigt, dass die Mehrheit der UserInnen von sozialen Netzwerken Hate Speeches schon einmal untergekommen sind. Jedoch werden diese zu meist nicht gemeldet, sondern ignoriert.

In den letzten Jahren gab es immer wieder mediale Berichterstattungen über Anklagen wegen Hasspostings. Vor allem der Fall von Eva Glawischnig hat hohe Wellen geschlagen, da sie im Kampf gegen Hasspostings erstmals erfolgreich gegen Facebook vorging. Dieser Fall hat jedoch auch gezeigt, dass eine Regulierung sozialer Netzwerke bezüglich des Umgangs mit Hate Speeches dringend notwendig ist.

## 4 Die Regulierung von sozialen Netzwerken

Dieses Kapitel befasst sich hauptsächlich mit derzeitigen Regulierungen von sozialen Netzwerken in einigen EU-Ländern. Zudem werden die verschiedenen Regulierungsmöglichkeiten beschrieben. Abschließend wird auf die immer wieder auftauchende Problematik mit der Meinungsfreiheit eingegangen.

### 4.1 Haftung als Plattformbetreiber

Grundsätzlich gibt es keine medienrechtliche Regulierung, die für Netzwerke und Plattformen gilt, wenn deren Betreiber keine redaktionelle Verantwortlichkeit wahrnehmen beziehungsweise keinen Einfluss auf die verbreiteten Inhalte und die Gestaltung des Gesamtangebots haben. Vor allem gilt dies für Plattformen von NutzerInnen der Dienste wie Facebook oder YouTube. Die Internetplattformen selbst, die bei der Verbreitung von Inhalten mitwirken, können sich auf die Haftungsprivilegien des E-Commerce-Gesetzes beziehen. Vor allem kommt der Haftungsausschuss des Hosting-Providers zum Tragen, so dass der Provider dann zur Verantwortung gezogen werden kann, wenn die zur Kenntnis gelangten rechtswidrigen Informationen nicht sofort gelöscht oder der Zugang gesperrt wird. Wichtig zu erwähnen ist hier jedoch, dass laut des Mediengesetzes immer die NutzerInnen, die Videos gestalten oder Blogs betreiben als Medieninhaber angesehen werden, wenn diese für die Gestaltung verantwortlich sind und die Verbreitung veranlassen. Dies gilt auch bei Betreiber moderierter Webseiten. Daher sind für Profile auf Facebook auch die ProfilbetreiberInnen medienrechtlich verantwortlich. Das heißt für den Inhalt haften laut Mediengesetz immer die UserInnen. Diese Haftung der MedieninhaberInnen sind jedoch eingeschränkt, sobald es sich um Inhalte Dritter handelt, die auf einer bestimmten Webseite bereitgestellt werden und die MedieninhaberInnen die nicht gebotene Sorgfalt dieser Seite ignoriert haben. Hauptsächlich bezieht sich diese Regelung auf Postings von dritten Personen, wie diese beispielsweise in der Form von Online-Leserbriefen oder Eintragungen auf Online-Diskussionsforen vorgenommen werden. (vgl. Berka & Trappel, 2019, S. 14)

Dadurch, dass Plattformbetreiber in der gesellschaftlichen Kommunikation und der Kommunikationskultur eine große Rolle spielen, wird häufig über eine Erweiterung

der ihnen auferlegten Haftung oder Erweiterung der Verpflichtung zur Löschung rechtswidriger Inhalte diskutiert. Die österreichische Politik erwartet sich hier beispielsweise eine Lösung auf europäischer Ebene. Die Europäische Kommission hingegen appelliert hier an eine entsprechende Selbstregulierung der Betreiber von Plattformen oder auch in Verbindung mit Verfahren der Ko-Regulierung. (vgl. ebd. ,S. 15)

Grundsätzlich wird von der Governance-Forschung sowieso von der Grundannahmen ausgegangen, dass ein Wandel von Government zu Governance stattfindet. Vor allem die Media Governance Forschung konzentriert sich auf die Modi von Selbst- und Ko-Regulierung in der Medienpolitik ohne zu unterschlagen, dass der Staat in der Institutionalisierung von nichtstaatlichen Ansätzen ein zentraler Akteur ist und auch bleibt. (vgl. Puppis, 2010, 134ff.) Demnach kennzeichnet Media Governance den Übergang von der klassischen Top-down Medienpolitik zur horizontalen Steuerung und den immer mehr werdenden Machtverlust des Gesetzgebers in der Formulierung, Durchsetzung, sowie Implementierung von Medienpolitik. (vgl. Meier, 2011, S. 43)

#### **4.1.1 Regulierungsformen**

Um die oben genannten Begrifflichkeiten zu verstehen, werden diese hier nun kurz erläutert. Bei den verschiedenen Begrifflichkeiten wie der Selbst- und Ko-Regulierung gibt es verschiedenen Definitionen. Hierbei stützt man sich jedoch auf die Interpretation von Latzer et al. (2002).

Zum einen gibt es die staatliche Regulierung. Im engeren Sinn umfasst dies den Mittelpunkt der staatlichen Institutionen: Legislative, Exekutive und Judikative. Hier ist hervorzuheben, dass hauptsächlich die Verwaltungsbehörden beziehungsweise Exekutive, also die staatlichen Institutionen im engeren Sinn, über die Kompetenz verfügen, Verordnungen zu erlassen. Zum anderen gibt es die staatliche Regulierung im weiteren Sinn. Hierbei sind in erster Linie jene Regulierungsaktivitäten zu verstehen, die zwar dem Staat zugerechnet werden, jedoch in Distanz zur öffentlichen Verwaltung stehen. Die Regulierungsaufgaben werden von privatrechtsförmigen Institutionen wie beispielsweise einer GmbH und von weisungsfreien Kollegialbehörden mit „richterlichem Einschlag“ getragen.

Solche Kollegialbehörden besitzen meist einen gerichtsähnlichen Status und können zu meist agieren ohne von der Politik beeinflusst zu werden. Im Gegensatz dazu sind bei den privatrechtsförmigen Institutionen politisch motivierte Weisungen möglich. (vgl. Latzer et al., 2002, S. 43ff.) Grundsätzlich geht es bei der staatlichen Regulierung also um die Regulierung mittels Gesetze und Verordnungen, öffentlichen Leistungen oder Gebühren und Steuern. (vgl. Saurwein, 2019)

Im Unterschied dazu gibt es die Ko-Regulierung, bei der die Institutionen nicht hoheitlich tätig sind. Allerdings verfügen sie über eine ausdrücklich einseitige Grundlage und es herrscht ebenfalls eine starke staatliche Involvierung, zum Beispiel durch die staatliche Aufsicht über die Regulierungstätigkeit durch Struktur-, Ziel- oder Transparenzvorgaben. (vgl. Latzer et al., 2002, S. 46) Man spricht hier also von einem Mittel der Selbstregulierung auf rechtlicher Grundlage (Saurwein, 2019).

Bei der Selbstregulierung hingegen fehlen die rechtlichen Steuerungsressourcen komplett. Der Staat gibt hier keine genauen politischen, rechtlichen oder organisatorischen Vorgaben. Staatlichen Akteure ist es jedoch möglich unter Rückgriff auf alternative Steuerungsressourcen auf Selbstregulierungsprozesse einzuwirken. Aus diesem Grund wird hier zwischen Selbstregulierung im engen und weiten Sinne differenziert. Bei der Selbstregulierung im weiten Sinn haben staatliche Institutionen Einfluss auf Selbstregulierungsarrangements durch finanzielle Beiträge, personelle Involvierung oder Verträge. Das heißt, dass hier eine gewisse staatliche Involvierung vorliegt und sich diese auf die Selbstregulierungspraxis auswirken kann. Bei der Selbstregulierung im engen Sinn gibt es einen Zusammenschluss privater AkteurInnen zum Zweck eines gemeinsamen Zielerreichen und zwar ohne staatliche Involvierung. (vgl. Latzer et al., 2002, S. 46f.)

Neben diesen genannten Regulierungsformen gibt es jedoch auch die individuelle Selbstorganisation, die von dem weiten Begriff der Selbstregulierung umfasst wird. Hierbei handelt es sich um die individuelle Organisation privater Akteure ohne jegliche Einflüsse. (vgl. ebd., S. 10) Des Weiteren gibt es die sogenannte Marktsteuerung, bei der davon ausgegangen wird, dass Probleme individuell durch das Konsumverhalten gelöst werden. (vgl. ebd., S. 47) Beispielsweise ist

hier gemeint, dass NutzerInnen auf ein Produkt verzichten oder zu einem anderen wechseln. (Saurwein, 2019)

## 4.2 Der Verhaltenskodex der EU

Wie aus den vorherigen Kapiteln hervorgeht ist eine Regulierung der sozialen Netzwerke zur Vermeidung von Hate Speeches dringend notwendig. Jedoch kann nicht nur eine Regulierung hilfreich sein, sondern auch UserInnen können etwas dagegen tun. Natürlich wäre es einfach, wenn auf Hass im Netz mit noch mehr Hass reagiert werden würde. Häufig ist dies sogar die naheliegendste Reaktion, wenn man beschimpft wird oder ähnliches. Bei Betrachtung der sozialen Medien so zeigt sich, dass Aggression noch mehr Aggression hervorruft. Mittelfristig gesehen ist dies jedoch keine gute Idee, da diese das Klima vergiften. Daher kann es auch helfen, wenn einzelne Wortmeldungen andere dazu bringen können einer Mehrheitsmeinung zu widersprechen. Es kann auch helfen auf DiskussionspartnerInnen mit Respekt zu antworten, so dass diese möglicherweise umgänglicher werden. Ebenfalls kann es hilfreich sein, unfaire Diskussionsstile zu benennen. Das heißt, bei der Bekämpfung von Hate Speeches, hilft es als UserIn nicht einfach wegzuschauen, sondern sich aktiv dagegen vorzugehen. (vgl. Brodnig, 2016, S. 184ff.)

Wie schon erwähnt sind jedoch auch staatliche Regulierungen der sozialen Netzwerke notwendig, um Hate Speeches so gut es geht zu vermeiden. Obwohl Artikel 15 der E-Commerce Richtlinie der EU aus dem Jahr 2000 schon vorsieht, dass illegale Hasskommentare gelöscht werden müssen, sobald diese gemeldet werden (vgl. Europäisches Parlament, 2000, S. 13), hat sich die EU-Kommission im Jahr 2016 mit Facebook, Microsoft, Twitter und Youtube auf einen Verhaltenskodex geeinigt um zukünftig Hate Speeches entgegenzuwirken. Auch Instagram, Google+, Snapchat und Dailymotion haben diesem Kodex im Jahr 2019 zugestimmt. (vgl. European Commission, 2019) Mit der Unterzeichnung dieses Verhaltenskodex haben sich diese Unternehmen dazu verpflichtet illegale Hassreden innerhalb von 24 Stunden nach deren Bekanntwerden zu löschen. Die Unternehmen haben sich hierbei immer an die nationalen Rechte zu halten. Das heißt bei der Einschätzung, ob eine Hassrede illegal ist oder nicht, gilt das

nationale Recht. Mit welchen Methoden diese Unternehmen hierbei vorgehen, ist ihnen selber überlassen. Des Weiteren sind sie dazu verpflichtet die NutzerInnen in den Richtlinien darüber zu informieren, welche Art von Content auf deren Plattform nicht erlaubt ist. (vgl. ebd.)

Im Rahmen dieses Verhaltenskodex haben zwölf Organisationen in neun EU-Staaten (Österreich, Belgien, Dänemark, Frankreich, Deutschland, Italien, Niederlande, Spanien und Großbritannien) im Jahr 2016 ein sogenanntes Monitoring bezüglich illegaler Online-Hetze vorgenommen. Dabei kam heraus, dass am meisten Benachrichtigungen bezüglich illegaler Hassreden, nämlich 270, von Facebook kamen. 163 kamen von Twitter und 123 von YouTube. Microsoft hatte bis dahin noch gar keine illegalen Hassreden gemeldet. Wenn es um die Löschquote geht wurden bei Facebook 28,3% der gemeldeten Fälle gelöscht, bei Twitter 19,1% und bei YouTube 48,5%. (vgl. Jourová, 2016, S. 3)

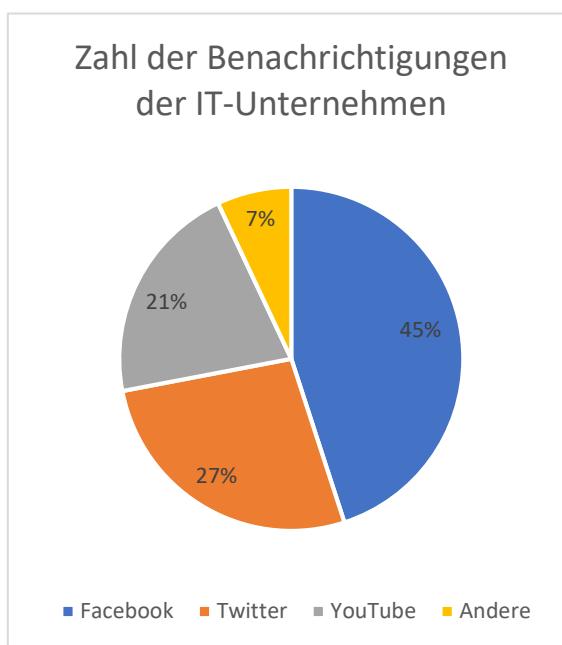


Abbildung 6: Zahl der Benachrichtigungen der IT-Unternehmen – Eigene Darstellung nach Jourová 2016, S. 3



Abbildung 7: % der gelöschten Inhalte aller Benachrichtigungen - Eigene Darstellung nach Jourová 2016, S. 4

Dieses Monitoring hat also ergeben, dass IT-Unternehmen den mit der Europäischen Kommission vereinbarten Umgang mit illegalen Hate Speeches noch unzureichend umsetzen. In Österreich lag die Löschquote mit 11% weit unter dem Durchschnitt, der bei 28% lag. Zum Vergleich: in Frankreich lag die Löschquote bei 58,1%, in Deutschland bei 52%, in Italien allerdings nur bei 3,6%.

Daher wurden die Unternehmen nach diesem Monitoring dringend dazu aufgerufen ihren Verpflichtungen nachzukommen und diese Vereinbarungen des Verhaltenskodex einzuhalten. (vgl. Demokratiezentrum Wien, 2019)

Seit 2016 wurde jährlich so ein Monitoring zur Kontrolle durchgeführt und es zeigt sich, dass die Benachrichtigungsrate illegaler Hassreden der IT-Unternehmen jährlich gestiegen ist.

**Rate of notifications reviewed within 24 hours since the launch of the Code of Conduct**

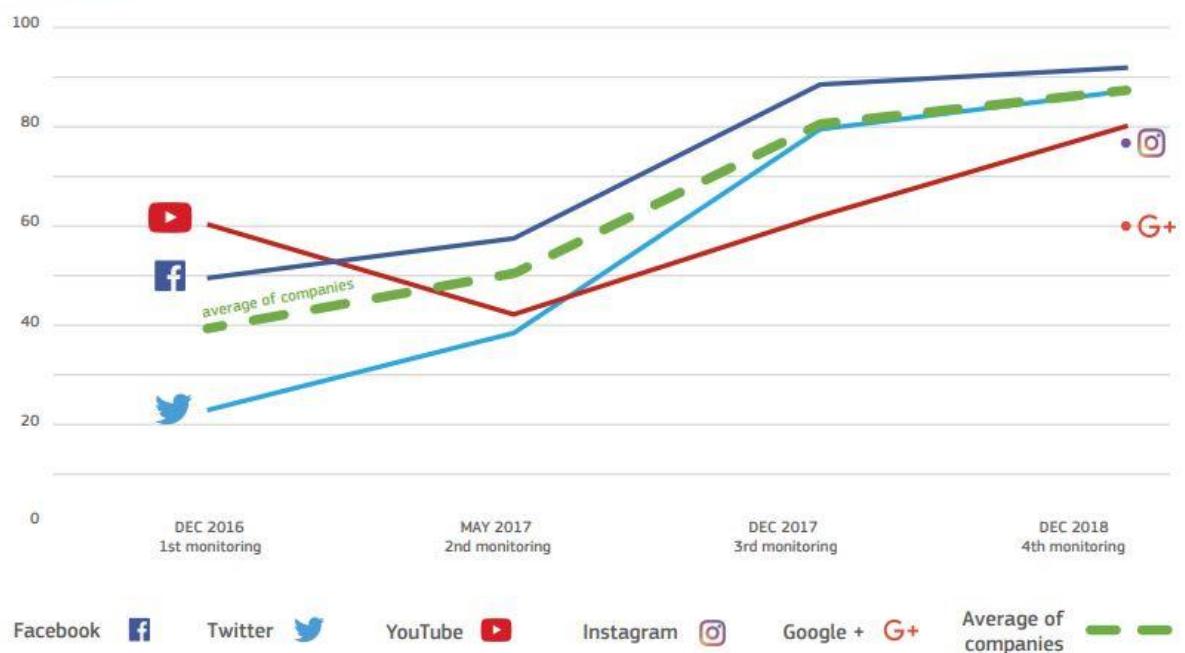


Abbildung 8: Rate of notifications reviewed within 24 hours since the launch of the Code of Conduct (Jourová 2019, S. 2)

Beim letzten Monitoring im Herbst 2018 haben sich 39 Organisationen von 26 EU-Mitgliedsstaaten beteiligt. Dabei kam heraus, dass die IT-Unternehmen 71,7% der gemeldeten Fälle gelöscht haben. Bei YouTube wurde von den gemeldeten Fällen 85,4% gelöscht, bei Facebook 82,4% und bei Twitter 43,5%. Die Löschquote hat sich also deutlich erhöht. (vgl. Jourová, 2019, S. 3)

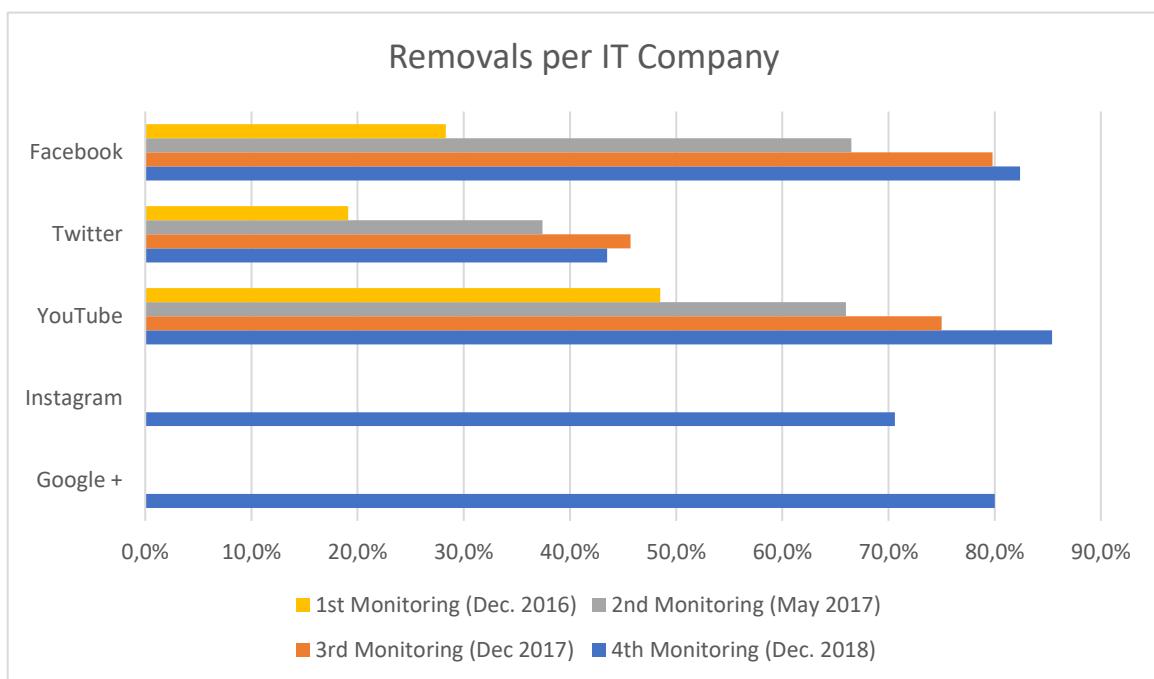


Abbildung 9: Removals per IT Company – Eigene Darstellung nach Jourova 2019, S. 3

#### 4.2.1 Wie soziale Medien gegen Hass im Netz vorgehen

Wie genau soziale Netzwerke gegen Hassreden im Netz vorgehen ist öffentlich nicht detailliert bekannt. Was man jedoch weiß ist, dass einige soziale Netzwerke über sogenannte Löschzentren verfügen. Das von Facebook hat beispielsweise seinen Sitz in der deutschen Stadt Essen, in dem 750 MitarbeiterInnen Jagd auf all das machen, was nicht auf der Plattform stehen soll: Hass, Gewalt und Propaganda. Die LöscharbeiterInnen sind quasi dafür zuständig den Müll rauszubringen, den Millionen NutzerInnen weltweit täglich auf Facebook abladen. (vgl. Stegler, 2018)

Grundsätzlich ist es so, dass Facebook-UserInnen gewisse Inhalte an Facebook melden. Diese Inhalte müssen daraufhin von Content ModeratorInnen des Löschzentrums gesichtet und gegebenenfalls entfernt werden. Die MitarbeiterInnen machen sich aber nicht proaktiv auf die Suche nach Content, der gegen die Regeln verstößt. Was unter Hassreden fällt, definiert sich nach den sogenannten Community-Standards. Daher ist es eben auch die Aufgabe dieser MitarbeiterInnen zu prüfen, wie diese Regeln in Bezug auf einen jeweils speziellen

Post auszulegen sind. Wie genau diese Löscharbeiten jedoch ablaufen, ist öffentlich kaum bekannt. (vgl. Hoppenstedt, 2017)

Viele der sozialen Netzwerke setzen bei der Bekämpfung von Hassreden aber auch auf die künstliche Intelligenz. Bei Facebook analysiert KI bereits alle auf den hauseigenen Plattformen veröffentlichte Inhalte, die gegen die Richtlinien verstößen, automatisch und meistens ist die Erfolgsrate auch sehr hoch. Facebook möchte nun sogar noch mehr in diese Technologie investieren um Inhalte künftig mit so wenig menschlicher Überprüfung wie möglich zu verstehen. (vgl. Rondinella, 2019) Auch Twitter verwendet die Künstliche Intelligenz zur Erkennung von Hate Speech. (vgl. Kleine Zeitung, 2018) Das Plattformunternehmen hatte 2017 angekündigt, künftig mittels Algorithmus gehen Hate Speeches vorzugehen. Accounts, die beleidigende Inhalte verbreiten, werden nun per Algorithmus erkannt. Hate Speeches werden also noch vor der Meldung von NutzerInnen erkannt. Des Weiteren möchte Twitter transparenter werden und die einzelnen Schritte nach einer Meldung einer Hassreden nachvollziehbarer machen. (vgl. Mühlenmeier, 2017) YouTube verwendet ein erst kürzlich entwickeltes System: Conversation AI. Dabei handelt es sich um sogenannte Sprach-Filter, die gegen Hassreden vorgehen sollen. In einem Test der Software stellte sich jedoch fest, dass das System auch vulgäre, aber freundliche Worte als gefährlich einstuft und löscht. (vgl. Süddeutsche Zeitung, 2018)

Um noch effektiver gegen Hassreden vorzugehen möchte Facebook nun in diesem Sommer ein neues unabhängiges Aufsichtsgremium einrichten. Dieses soll aus 40 weltweiten Mitgliedern, zum Beispiel WissenschaftlerInnen, JuristInnen, JournalistInnen und andere ExpertInnen bestehen. Das Gremium soll künftige über strittige Fragen beraten, die Facebook-NutzerInnen ihnen vorschlagen. Beispielsweise ob Hasskommentare konsequent gelöscht werden sollen oder ob diese online bleiben, jedoch mit einer Art Warnhinweis versehen werden soll. Alle Entscheidungen, die dieses Gremium treffen, seien daraufhin auch für Facebook verbindlich. (vgl. Handelsblatt, 2019)

## 4.3 Nationale Gesetze in den EU-Staaten

Obwohl die Europäische Kommission eine Selbst- beziehungsweise Ko-Regulierung zur Bekämpfung von Hate Speeches vorsieht (vgl. Berka & Trappel, 2019, S. 15), wird in manchen EU-Staaten die Meinung vertreten, dass staatliche Maßnahmen dafür notwendig sind. Aus diesem Grund haben Länder wie Deutschland oder Österreich mittlerweile ein Gesetz beschlossen oder einen Gesetzestext entworfen, um gegen Hate Speeches in den sozialen Netzwerken vorzugehen.

In Österreich fallen öffentliche Beiträge, die zu Gewalt oder Hass gegen Personen aufgrund ihrer Religionszugehörigkeit, Nationalität, ethnischer Zugehörigkeit, Weltanschauung beziehungsweise sexueller Orientierung oder Hautfarbe aufstachelt, unter Verhetzung nach §283 StGB und macht sich somit strafbar. Die gilt für das „Offline“-Leben sowie auch für Äußerungen online. (vgl. Saferinternet.at, 2016) Diese ähnlichen Gesetzgebungen der EU-Mitgliedstaaten haben die Gemeinsamkeit, dass Fälle bezüglich sozialer Medien im Rahmen allgemeiner Gesetze behandelt werden können. Schwierig ist jedoch die Identifizierung diskriminierender Inhalte im Internet beziehungsweise in sozialen Medien, weil diese Gesetze zu meist vor der Entwicklung des Internets entstanden sind. (vgl. Deutscher Bundestag, 2018)

Obwohl die Rechtslage klar ist, stellt sich die Durchsetzung für die Online-Welt als schwierig dar. Daher haben einige Länder eigene Gesetze verabschiedet, um Hassreden im Internet entgegenzuwirken.

### 4.3.1 Österreich: Aufhebung der Anonymität

Um effektiver gegen illegale Inhalte im Internet vorzugehen, wolle man in Österreich nun das neue „Gesetz für Sorgfalt und Verantwortung im Netz“ einführen. Das heißt, wer zukünftig einen Beitrag im Internet verfassen möchte, kann dies nicht mehr anonym tun. Der Regierung zu Folge müssen NutzerInnen den Plattformbetreibern in Zukunft ihren Vornamen, Nachnamen und ihre Adresse angeben. Die Plattformunternehmen müssen dafür sorgen, dass diese Angaben echt sind. Wie sie das machen, bleibt den Unternehmen selbst überlassen. Sollte

es zu Ermittlungen kommen, so sind die Betreiber verpflichtet diese Daten an Behörden weiterzugeben. Obwohl die Plattformbetreiber über den vollständigen Namen und die Adresse der NutzerInnen verfügen, können diese öffentlich weiterhin Pseudonyme verwenden. Sollten sich Plattformen nicht an die Vorgaben dieses Gesetzes halten, sind hohe Bußgelder angedacht. Die KommAustria, die als Aufsichtsbehörde fungieren soll, kann bei schweren Vergehen eine Strafe bis zu 500.000 Euro verhängen. (vgl. Der Standard, 2019b) Ziel des Gesetzes soll es sein, die Verfolgung von Rechtsansprüchen im Falle rechtswidriger Postings zu erleichtern und den respektvollen Umgang der Poster in Online-Foren zu fördern. (Österreichisches Parlament, 2019)

Diese Regelung gilt jedoch nicht für alle Plattformen. Es sind nur jene betroffen, die mehr als 100.000 registrierte NutzerInnen haben, mehr Umsatz als 500.000 Euro im Jahr erzielen oder eine Presseförderung von mehr als 50.000 Euro jährlich erhalten. Außerdem sind Onlineverkaufsplattformen explizit von dieser Regelung ausgenommen, reine Bewertungsplattformen hingegen nicht. Durch diese Vorgaben sind jedoch auf Seiten, wie beispielsweise das FPÖ-nahe Portal unzensuriert.at, ausgenommen, das häufig dafür kritisiert werden, Hasspostings stehen zu lassen. Dies wird damit begründet, dass kleine Plattformen sowie Start-ups, die sich solche Regelungen möglicherweise nicht leisten können, nicht belastet werden. (vgl. Der Standard, 2019b)

In Kraft treten soll dieser Regelung im Jahr 2020. Zuvor müsse geprüft werden, ob das Gesetz mit dem Urteil des Europäischen Gerichtshofs 2016 gegen eine Vorratsdatenspeicherung vereinbart wäre. Außerdem ist noch unklar, ob dies einen Verstoß gegen die E-Commerce-Richtlinie der EU darstellt. Dienstanbieter im Netz müssen sich nämlich bloß an das Recht ihres Herkunftslandes halten. Das heißt Österreich dürfe keine strengeren Regeln auflegen als das Herkunftsland eines Betreibers. (vgl. ebd.)

Dieses noch nicht beschlossene Gesetz stößt jedoch jetzt schon auf Kritik, da dies eine schlechte und unverhältnismäßige Idee darstelle und man dadurch den Grundsätzen des gemeinsamen Binnenmarktes widersprechen, sowie die Digitalisierung in Österreich senken würde. Zudem wird kritisiert, dass dies das falsche Mittel zur Bekämpfung von Hass im Netz sei. (vgl. Fanta, 2019)

#### **4.3.2 Deutschland: Netzwerkdurchsetzungsgesetz**

Im Jahr 2017 hat Deutschland ein Gesetz erlassen, das gewisse Regeln für Anbieter sozialer Netzwerke bezüglich des Umgangs mit NutzerInnen-Beschwerden über Hasskriminalität und anderen strafbaren Inhalten im Netz und eine vierteljährliche Berichtspflicht der Anbieter vorgibt. Des Weiteren haben Opfer von Persönlichkeitsverletzungen im Internet somit einen Anspruch auf Auskunft über Bestandsdaten des Verletzers/der Verletzerin. Das Gesetz richtet sich vor allem an soziale Netzwerke wie Facebook, YouTube und Twitter, denen dadurch einige Pflichten auferlegt werden. Unter anderem eine Berichtspflicht, das Einrichten eines Beschwerdemanagements und die Pflicht zur Benennung von Zustellungsbevollmächtigten sowie Empfangsberechtigten. Bei dem Beschwerdemanagement geht es hauptsächlich darum, dass soziale Netzwerke dazu verpflichtet sind, illegale Inhalte hinsichtlich des Netzwerkdurchsetzungsgesetzes nach Kenntnis zu überprüfen und zu entfernen. Anzuwenden ist dieses Gesetz bei Betreibern von sozialen Netzwerken, die im Inland mehr als zwei Millionen registrierte NutzerInnen haben, jedoch nicht nur E-Mail- und Messenger-Dienste. Berufliche Netzwerke, Fachportale, Online-Spiele sowie Verkaufsplattformen sind von diesem Gesetz ebenfalls ausgenommen. Welche Inhalte als illegal gelten sind im §1 Absatz 3 des Netzwerkdurchsetzungsgesetzes aufgezählt. Unter anderem fallen darunter die Gefährdung des demokratischen Rechtsstaates, Volksverhetzung, Bildung krimineller Vereinigungen, Beschimpfung von Bekenntnissen, Religionsgesellschaften und Weltanschauungsvereinigungen, üble Nachrede, Verleumdung etc. Im Rahmen der Berichtspflicht müssen soziale Netzwerke, die im Kalenderjahr mehr als 100 Beschwerden über rechtswidrige Inhalte erhalten haben, einen deutschsprachigen Rechenschaftsbericht verfassen. Dabei muss über den Umgang mit solchen Beschwerden berichtet werden und daraufhin im Bundesanzeiger sowie auf der eigenen Homepage veröffentlicht werden. (vgl. Deutscher Bundestag, 2017)

Grundsätzlich ist es also so, dass soziale Netzwerke „offensichtlich rechtswidrige Inhalte“ löschen müssen. Dies muss innerhalb von 24 Stunden nach Eingang der Beschwerde passieren. Bei komplexeren Fällen gibt es eine sogenannten Sieben-Tage-Regel, um in dieser Zeit über eine Löschung oder Sperrung eines Beitrags

entscheiden zu können. Unter bestimmten Voraussetzungen kann auch diese überschritten werden. Bei Unsicherheiten bezüglich des Kontextes eines Beitrages, sollen NutzerInnen die Möglichkeit haben Stellung dazu nehmen zu können. Außerdem können die sozialen Netzwerke ihre Entscheidung über nicht offensichtlich rechtswidrige Inhalte an eine Form der freiwilligen Selbstkontrolle abgeben. Man spricht hier von einer „anerkannten Einrichtung der regulierten Selbstregulierung“. Gegründet, ausgestattet und betrieben wird diese Einrichtung von den Unternehmen selber, muss jedoch staatlich zugelassen und vom Bundesamt für Justiz überwacht werden. (vgl. ebd.)

Daneben sollen Personen, die Opfer von illegalen Hassreden sind, gegen UrheberInnen solcher Inhalte vorgehen können. Daher soll es Anbietern sozialer Netzwerke erlaubt sein in Einzelfällen Auskunft über vorhandene Bestandsdaten von UrheberInnen zu geben. Dafür ist jedoch eine gerichtliche Anordnung erforderlich. Soziale Netzwerke sind ebenfalls dazu verpflichtet AnsprechpartnerInnen in Deutschland für Justiz, StrafverfolgerInnen, Bußgeldbehörden und BürgerInnen zu benennen, an den sich die genannten Institutionen und Personen wenden können. Dieser sogenannte Zustellungsbevollmächtigte muss binnen 48 Stunden auf Anfragen und Beschwerden reagieren. Kommt es zu Verstößen dieses Gesetzes droht eine Bußgeldstrafe bis zu fünf Millionen Euro. (vgl. ebd.)

Wegen mangelnder Transparenz bezüglich Hasskommentaren soll Facebook nun solch ein Bußgeld in Höhe von zwei Millionen Euro zahlen. Das Bundesamt für Justiz erließ mittlerweile auch schon einen Bußgeldbescheid gegen das Unternehmen „Facebook Ireland Limited“, da bei der Veröffentlichung des Transparenzberichts über das erste Halbjahr 2018 gegen die Vorgaben des Netzwerkdurchsetzungsgesetzes verstoßen wurde. Diese Strafe ist vor allem damit begründet, da in dem Bericht von Facebook die Zahl der eingegangenen Beschwerden über rechtswidrige Inhalte unvollständig angegeben sei. Durch diese unvollständige Angabe würde in der Gesellschaft ein verzerrtes Bild über das Ausmaß solcher rechtswidrigen Inhalte sowie über die Art und Weise, wie das soziale Netzwerk damit umgehe, entstehen. (vgl. Roßmann, 2019) Im Vergleich zu YouTube, das für denselben Zeitraum 215.000 Meldungen auswies und Twitter

mit 265.000 Meldungen, gab es bei Facebook lediglich 886 Beschwerden, bei denen 1.704 Inhalte bemängelt wurden. (vgl. ebd.)

#### **4.3.3 Lettland: „Law On Information Society Services“**

Seit 2011 gibt es auch in Lettland ein Gesetz, dass zur Regulierung von Hassreden im Internet beitragen soll. Das sogenannte „Law On Information Society Services“ reguliert die Verbreitung von Hassreden sowie die Anstiftung zu Gewalt in sozialen Netzwerken. Es wird dabei vorgesehen, dass Dienstanbieter durch Aufsichtsbehörden angewiesen werden können Informationen herauszugeben oder bestimmte Aktivitäten zu verbieten, sofern dies für die Verhütung und Untersuchung von Straftaten und für die Einleitung eines Strafverfahrens notwendig ist. Das Gesetz soll vor allem auch der Durchsetzung des im Strafgesetzbuch geregelten „Triggerin of National, Ethic and Racial Hatred“ gelten. Im Gegensatz zu dem Netzwerkdurchsetzungsgesetz in Deutschland, bei dem die Überprüfungspflicht den sozialen Netzwerken selber zugewiesen ist, erfolgt diese hier durch externe Aufsichtsorgane. (vgl. Deutscher Bundestag, 2018)

#### **4.3.4 Großbritannien: Unabhängiges Aufsichtsorgan**

Momentan existiert in Großbritannien noch kein Gesetz, das sich ins besondere mit der Rechtsdurchsetzung hinsichtlich der Verbreitung von Hass und der Anstiftung zu Gewalt in sozialen Netzwerken beschäftigt. Immer wieder hat die britische Regierung wiederholt, dass das was offline illegal ist auch online illegal sei. (vgl. Deutscher Bundestag, 2018)

Nun wurde jedoch bekannt, dass Großbritannien Inhalte im Internet strenger regulieren möchte. Dafür soll eine neue unabhängige Regulierungsinstanz geschaffen werden, die für Internetfirmen einen Verhaltenskodex vorgibt. Wird gegen diesen Kodex verstoßen, so drohen Bußen gegen das jeweilige Unternehmen. Es kann jedoch auch zu Strafen gegen einzelne ManagerInnen kommen. Im Fokus steht dabei der Kampf gegen die Verbreitung von illegalen Inhalten wie beispielsweise Kinderpornografie oder auch Mobbing-Botschaften. Diese Regulierung soll jedoch nicht nur für soziale Netzwerke gelten, sondern auch für Online-Unternehmen, die ebenfalls in irgendeiner Art und Wiese Inhalte

über das Internet zur Ansicht anbieten oder als Vermittlungsplattform dienen. Finanziert soll dieser Regulierungsinstanz von der Branche werden. Diese Behörde wird in Großbritannien tätige Internetunternehmen überwachen und Strafen selbst verhängen dürfen. (vgl. Tribe, 2019)

Dies sind vier europäische Länder mit vier verschiedenen Regulierungsansätzen hinsichtlich Hate Speech. Über ähnliche Regulierungen in anderen EU-Mitgliedsstaaten ist bislang nichts bekannt. Italien hat jedoch erst vor Kurzem die Anwendung des Artikel 416 des italienischen Strafgesetzbuches auch Hassreden durch virtuelle Gemeinschaften, Blogs und soziale Netzwerke ausgeweitet und auch in Frankreich, hat der Premierminister Edouard Philippe ein Gesetz angekündigt, dass sich speziell gegen rechtswidrige Inhalte im Internet richten soll. Es soll dabei nach deutschem Vorbild vorgegangen werden, wo Anbieter von sozialen Netzwerken mit hohen Geldbußen rechnen müssen, werden Hate Speeches nicht innerhalb von 24 Stunden gelöscht beziehungsweise das Benutzerkonto der HassrednerInnen entfernt. (vgl. Deutscher Bundestag, 2018)

#### **4.4 Die Problematik mit der Einschränkung der Meinungsfreiheit**

Wenn es um das Thema Hate Speech geht fällt oft auch das Wort Meinungsfreiheit. Jedoch ist zu beachten, dass die Regeln des menschlichen Zusammenlebens bei der Online-Kommunikation nicht außer Kraft treten und die Meinungsfreiheit so auch nicht jede Wortmeldung erlaubt. Artikel 1 des deutschen Grundgesetzes lautet, dass die Würde des Menschen unantastbar sei. Diese ist zu achten und zu schützen ist die Verpflichtung aller staatlichen Gewalten. Es gibt hier keine Ausnahme, dass es im Internet anders sei. Die Meinungsfreiheit reicht jedoch nur so weit, bis die Rechte anderer verletzt werden. Beispielsweise ist damit das Recht gemeint, dass sich Personen nicht bedrohen oder verunglimpfen lassen müssen. Rechtlich gesehen ist es also falsch zu behaupten, dass die Meinungsfreiheit unbegrenzt sei. Im Zweifelsfall liegt es an den Gerichten abzuwägen, ob das Recht der freien Rede oder beispielsweise der Schutz vor übler Nachrede schwerer zu gewichten ist. Sich ständig auf die Meinungsfreiheit zu stützen ist hier vielmehr eine Immunisierungsstrategie: Es wird damit versucht,

für die eigenen Worte nicht gerade stehen zu müssen. (vgl. Brodnig, 2016, S. 86ff.)

Viele HassrednerInnen fühlen sich jedoch in Ihrer Meinungsfreiheit angegriffen, wenn viele UserInnen dagegenreden. Dies führt jedoch zu einem wesentlichen Aspekt der Meinungsfreiheit, denn diese erlaubt auch die Widerrede. Wenn jemand also furchtbare Aussagen über eine Personengruppe tätigt, wird die Meinungsfreiheit von demjenigen nicht eingeschränkt, nur weil andere BürgerInnen demjenigen ordentlich die Meinung sagen. Das ist sogar ihr gutes Recht, denn auch für die gilt die Meinungsfreiheit. (vgl. ebd., S. 92)

Für HassrednerInnen gilt also die Meinungsfreiheit genauso wie für die „besorgten BürgerInnen“, die dagegenreden. Jedoch eben nur so weit, bis Rechte anderer verletzt werden. (vgl. ebd., S. 90)

Einige PolitikerInnen sind dennoch besorgt. Sie sind weiterhin der Meinung, dass es sich beim dem Verhaltenskodex der EU um Zensur handle. Hingegen wäre es viel effektiver mit Aussagen, mit denen man nicht einverstanden sei, mit anderen Aussagen zu kontern. Vor allem ist man der Meinung, dass es kein allgemein gültiges Konzept der Hate Speech gibt. Das heißt was für eine Person als Hate Speech gilt, ist auch Sicht einer anderen Person überhaupt nicht anstößig. (vgl. Fox, 2018)

Es geht jedoch nicht nur um die, die den Hass verbreiten, sondern auch um die anderen UserInnen. Denn einige Regulierungsversuche der sozialen Netzwerke scheinen den Eindruck zu erwecken, dass dadurch die Meinungsfreiheit eingeschränkt werden würde. So zum Beispiel bei dem Netzwerkdurchsetzungsgesetz. Seit der Einführung dieses Gesetzes in Deutschland gibt es heftige Diskussionen rund um die Meinungsfreiheit. Zwar haben sich die Wogen mittlerweile etwas geglättet, jedoch befürchtet die internationale Menschenrechtskonvention, dass dieses Gesetz zur Zensur einlädt. Dies ist vor allem deshalb der Fall, da bei diesem Gesetz die Angst bestehe, dass zu viel und somit auch Beiträge, die keine Hate Speeches darstellen, gelöscht werden. (vgl. Kurz, 2018)

## 4.5 Zusammenfassung

Im Jahr 2016 hat die EU einen Verhaltenskodex für soziale Netzwerke ins Leben gerufen wurde. Dadurch wurde innerhalb der Gesellschaft immer mehr auf das Thema Hate Speech im Internet Aufmerksam gemacht. Daher versuchen Facebook, Twitter & Co. sich so gut es geht an den Kodex zu halten. Obwohl dies zu Beginn von den sozialen Netzwerken noch nicht sehr ernst genommen wurde, bemühten sich diese jährlich mehr gegen Hassreden im Internet vorzugehen. Vor allem auch deshalb, da immer mehr EU-Mitgliedsstaaten, trotz des Verlangens einer Selbst- beziehungsweise Ko-Regulierung der sozialen Netzwerke, an eigenen Gesetzen für soziale Netzwerke zur Vermeidung von Hassreden arbeiten, die greifen sobald der Verhaltenskodex verletzt wird. Bislang hat hauptsächlich Deutschland ein effektives Gesetz für soziale Netzwerke gegen Hassreden erlassen, das vorsieht illegale Inhalte innerhalb 24 Stunden zu löschen. Wird dies nicht eingehalten, drohen hohen Geldstrafen. In anderen Ländern ist man weiterhin am Entwerfen eines eigenen Gesetzes. Wenn es also um die Regulierung geht, sieht man, dass es momentan in Richtung staatlicher Regulierung geht. Jedoch gibt es auch einige Gegner davon, Hate Speeches mittels nationaler Gesetze zu regulieren, da diese befürchten, dass dadurch die Meinungsfreiheit eingeschränkt werden würde. Dadurch wird häufig der Eindruck erweckt, als wäre eine Regulierung nicht notwendig.

## 5 Zwischenfazit & Forschungsfragen

Durch das Aufkommen sozialer Netzwerke hat sich die Kommunikation innerhalb der Gesellschaft wesentlich verändert. Während man vor der Zeit des Internets hauptsächlich persönliche Gespräche mit seinem Gegenüber führte, so geschieht dies nun über das Internet. Zudem hat sich auch die Art der Informationsbeschaffung und somit die Art Meinungsbildung verändert.

Innerhalb der sozialen Netzwerke wird häufig auf das Phänomen der Hate Speech gestoßen. Hierbei handelt es sich um öffentliche Botschaften mit diskriminierenden Inhalten. Betroffene dieser Hate Speech erleiden häufig einen seelischen Schaden. Jedoch hat es nicht nur einen Einfluss auf die Opfer, sondern auch auf die Meinungsbildung anderer. Gerät man nämlich in eine sogenannte Filterblase in den sozialen Netzwerken, in der es nur um die Verbreitung von Hass geht, so kann die Realität von UserInnen sehr schnell verzerrt und die Meinung eines Einzelnen beeinflusst werden. Des Weiteren kann eine starke Verbreitung von Hate Speech zu einer nicht-geordneten Gesellschaft führen, da Minderheiten in der Gesellschaft womöglich nicht mehr respektiert werden würden.

Sieht man sich Statistiken an, sind den meisten NutzerInnen schon einmal Hasspostings oder Hasskommentare in sozialen Netzwerken untergekommen. Vor allem auf Facebook sind Hassreden sehr verbreitet und richten sich zumeist gegen Personen mit Migrationshintergrund. Viele UserInnen reagieren jedoch nicht auf solche Hassreden in sozialen Netzwerken, sondern ignorieren diese.

Um mehr Aufmerksamkeit auf dieses Thema zu richten und Hassreden in sozialen Netzwerken zu mindern, hat die EU im Jahr 2016 einen Verhaltenskodex für soziale Netzwerke eingerichtet. Alle sozialen Netzwerke, die diesem Kodex zugestimmt haben, sind also verpflichtet, nach Bekanntwerden eines Hasspostings, diesen innerhalb von 24 Stunden zu löschen. Wie dabei vorgegangen wird bleibt den Unternehmen selber überlassen. Facebook arbeitet hierbei beispielsweise mit Löschzentren und der künstlichen Intelligenz. Auch YouTube wendet zur Erkennung von Hassreden die künstliche Intelligenz in Form von Sprachfiltern an.

Um die Regulierung sozialer Netzwerke zu verstärken, haben sich auch einige Länder dazu entschieden, ein Gesetz zu erlassen, das Hate Speeches verhindern

soll. In Deutschland wurde beispielsweise das Netzwerkdurchsetzungsgesetz erlassen, das vorsieht offensichtlich illegale Inhalte innerhalb von 24 Stunden zu löschen. Auch in Österreich möchte man mittels eines Gesetzes für mehr Sorgfalt und Verantwortung im Internet gegen Hassreden vorgehen.

Trotz dieser Versuche der staatlichen Regulierung sieht man hier die Meinungsfreiheit gefährdet. Vor allem hinsichtlich des Netzwerkdurchsetzungsgesetzes in Deutschland ist man der Meinung, dass es zu einem Overblocking kommt und so zu viele Inhalte gelöscht werden.

Obwohl heute versucht wird, effektiv gegen Hassreden in sozialen Netzwerken vorzugehen, werden viele angewandte Methoden in Frage gestellt. Aus diesem Grund ergibt sich folgende Leitfrage: „Wie können soziale Netzwerke zukünftig reguliert werden um Hate Speech zu vermeiden?“

Um diese Leitfrage genau beantworten zu können, wurden folgende Forschungsfragen gebildet.

**FF 1:** Welche Rolle spielen Techniken wie „Artificial Intelligence“ bei der Bekämpfung von Hassreden und in sozialen Netzwerken?

**FF 2:** Wie effizient ist eine Identifikationspflicht bzw. Ausweispflicht, die in Österreich für 2020 vorgesehen ist um gegen Hass im Netz vorzugehen?

**FF 3:** Wieso werden diese Angelegenheiten nicht auf EU-Ebene geregelt?

**FF 4:** Welche Rolle spielt die politische Situation in den einzelnen Ländern, wenn es um die Regulierung von sozialen Netzwerken geht?

## 6 Methodik

Der vorherige Abschnitt dieser Forschungsarbeit stellt die Kernbereiche der Thematik dar. Dabei geht es um die sozialen Netzwerke und die Problematik der Regulierung zur Vermeidung von Hate Speeches. Diese theoretische Ausarbeitung bildet die Basis um im folgenden Abschnitt die Leitfrage mittels empirischer Untersuchung beantworten zu können.

Das nächste Kapitel beschreibt die Methode, die für diese Forschung angewandt wurde. Dabei werden zuerst das Forschungsdesign und daraufhin die Auswahl der InterviewpartnerInnen und die genaue methodische Vorgehensweise der Datenerhebung sowie -auswertung erläutert.

### 6.1 Forschungsdesign

Um die Leitfrage dieser Forschungsarbeit „Wie können soziale Netzwerke zukünftig reguliert werden um Hate Speech zu vermeiden?“ ausreichend beantworten zu können ist eine empirische Untersuchung dieser Thematik notwendig. Dafür wurde ein Forschungsdesign entworfen, das ein grundlegendes Konzept der methodischen Vorgehensweise darstellen soll. Die Ausarbeitung dieses Konzepts erfolgt nach dem entsprechenden Forschungsinteresse.

Laut Mayring (2010) kann ein Forschungsdesign explorativ, deskriptiv oder kausal gestaltet werden. Bei den explorativen Studien geht es darum ein Thema näher zu präzisieren und erste Erkenntnisse zu gewinnen. Die deskriptiven Designs stellen die Beschreibung von Verhaltensmerkmalen und Zusammenhänge dar und bei den kausalen Designs wird versucht Ursachen- sowie Wirkungszusammenhänge zu erkennen.

Obwohl das Thema Hate Speech, sowie soziale Netzwerke, mittlerweile nichts Neues mehr sind, wird dieses Forschungsdesign explorativ erforscht. Das Ziel der explorativen Forschung ist es nämlich neue Erkenntnisse zu gewinnen und daraufhin das Forschungsgebiet präzisieren zu können. Idealerweise werden darauffolgend auch Hypothesen abgeleitet und formuliert, die als Grundlage für eine weitere quantitative Forschung dienen können. (vgl. ebd., S. 232)

Da es bisher noch keine Studien über die Regulierung sozialer Netzwerke zur Vermeidung von Hate Speeches gibt, stellt dies eine Forschungslücke dar, die durch diese Forschungsarbeit gefüllt werden soll.

Solch eine explorative Forschung wird zumeist, wie auch in dieser Arbeit, in Form von qualitativen Beobachtungen oder Interviews, insbesondere ExpertInneninterviews, durchgeführt. (vgl. ebd.)

## **6.2 Erläuterung der Erhebungsmethode**

Grundsätzlich herrscht in der Forschung Einigkeit, dass eine qualitative Forschung häufig mittels qualitativer Interviews durchgeführt wird. (vgl. Mey & Mruck, 2010) Daher wurde auch hier das ExpertInneninterview zur Beantwortung der Leitfrage herangezogen, das sich auf einen Leitfaden stützt.

### **6.2.1 Leitfadengestütztes ExpertInneninterview**

Bei dem ExpertInneninterview handelt es sich um eine Methode der qualitativen empirischen Sozialforschung, bei der das Wissen von ExpertInnen zur Beantwortung der Forschungsfragen verwendet wird (vgl. Meuser & Nagel, 2009, S. 35). Im Gegensatz zu den anderen Methoden, wie der Beobachtung, einer Feldstudie oder anderen systematischen quantitativen Untersuchungen, ermöglicht das ExpertInneninterview eine konkurrenzlos dichte Datengewinnung (vgl. Bogner & Menz, 2009, S. 7). Laut Flick (2007) steht bei dieser Art der Befragung weniger die befragte Person im Mittelpunkt, sondern viel mehr die Eigenschaft als ExpertIn für ein spezifisches Handlungsfeld. Genauer gesagt wird hierbei auf exklusives Wissen abgezielt. Es wird versucht Personen zu befragen, die etwas erzählen können, das noch nirgends aufgeschrieben wurde (vgl. Löblich, Meyen, Pfaff-Rüdiger, & Riesmeyer, 2011, S. 61). Es geht bei diesen qualitativen Interviews also weniger um den „Sinn“ als vielmehr um das Wissen der ExpertInnen. Daher ist die Reaktivität auch niedriger als bei normalen qualitativen Befragungen. (vgl. ebd., S. 62) Als das passendste Erhebungsinstrument bei dem ExpertInneninterview eignet sich das offene Leitfadeninterview (vgl. Meuser & Nagel, 2009, S. 52). Obwohl sich das Leitfadeninterview sehr wohl vom ExpertInneninterview unterscheidet (vgl. Löblich

et al., 2011, S. 61), wäre es ungünstig auf einen Leitfaden und somit auf eine thematische Vorstrukturierung zu verzichten. Ohne eine entsprechende Vorbereitung des Interviews besteht die Gefahr sich den ExpertInnen gegenüber als inkompetente/r GesprächspartnerIn darzustellen. (vgl. Meuser & Nagel, 2009, S. 52) Außerdem würde dies methodisch gesehen in die falsche Richtung gehen. Um also ein möglichst erfolgreiches Interview zu ermöglichen ist es essenziell viel Arbeit in die Erstellung eines Leitfadens zu investieren. Die Behauptung „Dummheit als Methode“ würde in die richtige Richtung führen, stellt sich als falsch heraus. Es ist wichtig im Vorhinein genug Informationen zu sammeln, sowie über aktuelle Geschehnisse, die unter anderem in der Presse veröffentlicht wurden, Bescheid zu wissen. Ist man für das Interview nicht ausreichend informiert, wirkt man als GesprächspartnerIn nicht nur inkompetent, sondern es lässt auch die Vermutung aufkommen, nicht interessiert zu sein. (vgl. ebd., S.52) Die Fragen, die während des Interviews gestellt werden, sollten sich auf das Wie des Entscheidens und des Handelns fokussieren. Durch die richtige Formulierung der Fragen soll verständlich gemacht werden, dass diese auf das institutionsbezogene Wissen abzielen. Die persönlichen Qualifizierungen werden sowieso erwähnt, es sei denn der/die Interviewte zieht sich auf seine formale Rolle zurück. (vgl. ebd., S. 54) Die Auswertung des ExpertInneninterviews richtet sich nach thematischen Einheiten an inhaltlich zusammengehörigen über die Texte verteilten Passagen. Der gemeinsam geteilte institutionell-organisatorische Kontext ist es am Ende der die Interviews vergleichbar macht. Zudem wird die Vergleichbarkeit ebenfalls durch die leitfadenorientierte Interviewführung garantiert. Der Leitfaden filtert die benötigten Themen aus dem Horizont möglicher Gesprächsthemen der ExpertInnen heraus um sich während des Interviews auf diese Themen zu fokussieren und um nicht von den wesentlichen Themen abzuschweifen. (vgl. ebd., S. 56)

### **6.2.2 Auswahl der ExpertInnen**

In der heutigen Gesellschaft spielen Expertise sowie ExpertInnenwissen eine große Rolle. Dies ist nicht nur in der Technik oder in der Wissenschaft von großer Bedeutung, sondern auch in Bereichen des alltäglichen Lebens. Häufig handeln Menschen heute nicht mehr nach ihrer Intuition, sondern anhand von

wissenschaftlicher Expertise. Als ExpertIn werden also jene Personen bezeichnet, die ein Sonderwissen in bestimmten Bereichen besitzen. Das ExpertInnen-Sein ist jedoch keine personelle Eigenschaft oder Fähigkeit, sondern vielmehr eine Zuschreibung. In der Praxis findet eine solche Zuschreibung statt sobald wir auf Grund unserer Forschungsinteressen mit ExpertInnen Kontakt aufnehmen. (vgl. Bogner et al., 2014, S. 10f.) Für ein ExpertInneninterview werden nur noch jene ExpertInnen befragt, die auch als solche wahrgenommen werden. Solch eine Bestimmung ist jedoch nicht ausreichend, da ForscherInnen keine Kriterien vorliegen nach denen sie sich richten können um ExpertInnen von Nicht-ExpertInnen unterscheiden zu können. Werden solche Kriterien nicht aufgestellt, so besteht die Gefahr, dass der Begriff der ExpertInnen soweit ausgedehnt wird, dass folglich jeder Person so ein Status zugesprochen wird. Aus diesem Grund ist es wichtig, dass bei der Bestimmung von ExpertInnen danach gefragt wird, welche Merkmale es gibt, die das Handeln sowie das Wissen der ExpertInnen von anderen unterscheidet. (vgl. Meuser & Nagel, 2009, S. 37)

Um also InterviewpartnerInnen auswählen zu können, muss im Vorhinein ausführlich darüber nachgedacht werden, wer über das notwendige Wissen verfügt um die Fragen während des Interviews auch beantworten zu können. Um alle Informationen zu sammeln, die zur Beantwortung der Forschungsfragen notwendig sind, müssen zumeist mehrere AkteurlInnen befragt werden. Am einfachsten wäre es, wenn man Personen innerhalb des Freundeskreises oder in der Familie kennen würde, die als ExpertInnen gelten würden. Diese sind nämlich zumeist am ehesten für ein Interview bereit. Davor sei jedoch abzuraten, da sonst die Gefahr besteht, dass diese sich als jemanden ausgeben, die sie gar nicht sind. Vor allem verändern bereits vorhandene und persönliche Beziehungen die Interviewsituation. (vgl. Gläser & Laudel, 2009, S. 117f.)

Um für diese empirische Untersuchung die richtigen ExpertInnen zu finden wurden Kriterien definiert, die bei der Auswahl berücksichtigt wurden. Die Grundgesamtheit dieser Forschungsarbeit umfasst alle Personen mit Unionsbürgerschaft der Europäischen Union, die entweder in der Forschung tätig sind und zu deren Forschungsgebiet Hate Speeches beziehungsweise Regulierung von Internet-Plattformen zählen oder Personen, die in diesem Bereich

beruflich tätig sind. Zudem werden auch JuristInnen dazugezählt, die sich unter anderem mit dem EU-Recht beschäftigen. Aus dieser Grundgesamtheit wurden sechs ExpertInnen ausgewählt. Die nachfolgende Tabelle gibt eine Übersicht über die ausgewählten ExpertInnen:

Fall	Name	Unternehmen/ Organisation	Position im Unternehmen/der Organisation	Forschungs- gebiet
LS	Liriam Sponholz	Ludwig-Maximilian-Universität / Institut für Kommunikationswissenschaften und Medienforschung	Privatdozentin	Hate Speech, Journalistische Objektivität, Mediensysteme in Westeuropa und Lateinamerika etc.
FS	Florian Saurwein	Institute for Comparative Media and Communication Studies	Senior Scientist	Media-Governance, Medienethik, etc.
MKT	Meike Kolck-Thudt	ZARA – Zivilcourage & Anti-Rassismus-Arbeit	Zuständig für: Medienarbeit, Presseanfragen, Veranstaltungen	-
AW	Andre Wolf	Mimikama – Verein zur Aufklärung von Internetmissbrauch	Kommunikations-experte & Pressesprecher	-
SR	Stephan Ruhmannseder	Demokratiezentrum Baden-Württemberg	Projektleitung der Meldestelle „respect!“	-
VF	Verena Fabris	No-Hate-Speech Komitee	Nationale Koordinatorin	-

Tabelle 4: Auflistung der ExpertInnen – Eigene Darstellung

### 6.2.3 Leitfaden

Wie vorhin erwähnt, handelt es sich bei Leitfadeninterviews im Rahmen eines ExpertInneninterviews um ein teilstrukturiertes Interview. Leitfäden für das Interview mit ExpertInnen haben zwei Funktionen: Zum einen sollen sie helfen das Themenfeld der Untersuchung in verschiedene Teile einzugliedern, zum anderen sollen sie als Hilfsmittel während der Erhebung dienen. Oft variiert die Länge des Leitfadens. Wie detailliert solch ein Leitfaden ist, hängt von dem Interviewstil der ForscherInnen ab. Anders als in der quantitativen Sozialforschung, bei der die Fragen standardisiert sein müssen, müssen bei den qualitativen Interviews nicht

unbedingt dieselben Fragen bei den Befragungen gestellt werden, um die Interviews später vergleichen zu können. Maßgeblich ist es, die Befragten zum Reden zu bringen und dabei soll der Leitfaden als Gedächtnisstütze dienen. Bei der Erstellung eines Leitfadens soll mit Themenblöcken gearbeitet werden. Bei einem ein- bis zweistündigen Experteninterview soll es beispielsweise drei bis acht Themenblöcke geben, für welche jeweils drei Hauptfragen erstellt werden sollen, die als Gesprächsanreiz dienen. Zusätzlich werden aber noch weitere abhängige Fragen erstellt, die dazu führen sollen, während des Interviews etwas ins Detail zu gehen. Während die Hauptfragen verpflichtend sind, werden die ergänzenden Fragen nur gestellt, wenn die Hauptfragen nicht detailliert genug beantwortet werden. Während des Interviews kommt es oft vor, dass einige Fragen des Leitfadens gar nicht mehr gestellt werden müssen, da diese während des Gesprächs eines anderen Themenblocks schon beantwortet wurden. Zudem kommt es manchmal vor, dass während des Interviews spontane Zwischenfragen entstehen, die sich nicht im Leitfaden befinden. Aus diesem Grund kann die konkrete Erhebungssituation nie genau vorhersagt werden (vgl. Bogner et al., 2014, S. 27f.) Häufig ist es auch so, dass die Chronologie der Fragen in der Realität eine andere ist als vorgesehen. Gehen die Befragten bei der Beantwortung einer Frage auf ein Thema ein, das eigentlich erst für später vorhergesehen ist, so sollte dieser Gesprächsrichtung nicht entgegengesteuert werden. Daher sollten die Themenblöcke so gestaltet sein, dass diese in sich abgeschlossen sind und als Gesprächsbausteine versetzt werden können. Flexibilität ist hier also sehr wichtig. (vgl. ebd. S. 29f.)

Im Rahmen dieser Forschungsarbeit wurde ein Leitfaden für die ExpertInneninterviews erstellt, der aus fünf Themenblöcken besteht. Bei der Erstellung dieser Themenblöcke wurde sich an der literarischen Ausarbeitung sowie den erarbeiteten Forschungsfragen orientiert. Im ersten Themenblock wurde das Thema „Social Media“ sowie der Einsatz von Artificial Intelligence zur Erkennung von Hate Speeches angesprochen. Daraufhin folgten zwei separate Themenblöcke, die sich jeweils mit Gesetzen zur Regulierung von Hate Speeches beschäftigten. Im vierten Themenblock geht es um die Regulierung auf EU-Ebene.

Abschließend folgte ein Abschnitt, der sich mit politischen Einflüssen bezüglich der Regulierung beschäftigte.

## **6.3 Umsetzung der Erhebungsmethode**

### **6.3.1 Pretest**

Laut Bogner, Littig und Menz (2014) ergeben sich bei der Erstellung des Leitfadens häufig die Fragen wie dies in der Praxis umsetzbar ist und ob der geplante Zeitrahmen auch angemessen ist. Außerdem erwähnen Gläser und Laudel (2009), dass es zu der Situation kommen kann, dass die Fragen des Leitfadens nicht das bewirken, was sie sollen. Daher ist es ratsam, eine oder mehrere Probeinterviews mit Personen durchzuführen, die eventuell die gleichen Merkmale haben, wie die späteren InterviewpartnerInnen, um den Leitfaden vor der eigentlichen Untersuchung zu testen. Dementsprechend wurde auch für diese Forschungsarbeit ein Pretest des Leitfadens durchgeführt. Das Interview fand mit einer Person statt, die zwar weder in der Forschung tätig ist, noch beruflich mit Hate Speeches zu tun hat, sich jedoch privat sehr häufig mit diesem Thema beschäftigt. Während der Durchführung dieses Pretests wurde ersichtlich, dass zwei Fragen nicht deutlich genug formuliert wurden. Außerdem fiel auf, dass bei der Einstiegsfrage schon sehr viele Unterfragen der verschiedenen Themenblöcke beantwortet werden können. Bezuglich des zeitlichen Rahmens haben 45 Minuten durchaus gereicht.

### **6.3.2 Interviewablauf**

Allgemein wird über die Beziehung zwischen InterviewerInnen und ExpertInnen gesagt, dass die InterviewpartnerInnen sowie jemand selbst zu respektieren sind. Während der Entwicklung einer Interviewbeziehung soll darüber nachgedacht werden, was jemandem selbst und den InterviewpartnerInnen gegenüber fair ist. Eine solche Entwicklung der Interviewbeziehung passiert meist in den persönlichen Kontakten vor dem Interview. Vor allem in der Anfangsphase, denn dort werden die Erwartungen an die InterviewpartnerInnen erstellt und präsentieren sich im selben Zeitpunkt als mit einem bestimmten Fachwissen und Professionalität ausgestattete WissenschaftlerInnen. Diese beiden Positionen

müssen ausbalanciert werden. Es ist äußerst wichtig sich den InterviewpartnerInnen gegenüber respektvoll zu verhalten und deren Expertise zu respektieren, aber zugleich auch selber ExpertInnenwissen zu zeigen. Die InterviewerInnen sollen in deren Rolle als ForscherInnen wahrgenommen und auch anerkannt werden. Die Vorbemerkungen sowie die ersten Interviewfragen sollen dabei helfen. Ist der Einstieg zu locker, so kann dies die professionelle Rolle zerstören und eine Fehl wahrnehmung der ExpertInnen erzeugen. Dadurch kann es passieren, dass die Interviewten während des gesamten Interviews immer nur kurze und allgemeine Antworten geben, anstatt in die erwünschte Tiefe zu gehen. Am wichtigsten ist es also, dass das Interview so natürlich wie möglich bleibt. (vgl. Gläser & Laudel, 2009, S. 172)

Neben all diesen allgemeinen Regeln zum Aufbau einer Beziehung im Interview beschreiben Gläser und Laudel (2009) sieben spezielle Verhaltensregeln:

1. Aktiv Zuhören. Dabei soll sich auf den Inhalt der Ausführungen der InterviewpartnerInnen konzentrieren werden, ihnen verständlich machen, dass es verstanden wird und einschätzen können, welche Informationen noch gebraucht werden. Wichtig ist hierbei auch, dass die Interviewten bei ihren Erzählungen nicht unterbrochen werden und dass Pausen zugelassen werden.
2. Während einer möglichst natürlichen Gesprächssituation sollte sich nicht davor gescheut werden flexibel Fragen zu stellen, denn die Antworten der InterviewpartnerInnen bieten häufig solche Möglichkeiten. Dadurch können zusätzliche Erkenntnisse gewonnen werden, die für die Beantwortung der Forschungsfragen ebenfalls relevant sein können.
3. Wird etwas in den Ausführungen der InterviewpartnerInnen nicht verstanden, so müssen diese um Erklärung gebeten werden. Damit wird deutlich gemacht, dass es für die InterviewerInnen interessant ist, das Gesagte zu verstehen.
4. Häufig wird angenommen, dass offene Fragen automatisch zu dem gewünschten detailreichen Erzählfluss führen. InterviewpartnerInnen neigen jedoch dazu allgemein, mehrdeutig und vage zu bleiben. Aus diesem Grund ist es wichtig Details zu erfragen.

5. Werden Nachfragen gestellt sollten diese kurz und eindeutig sein. Ein typischer Fehler ist es nämlich, mehrere Fragen hintereinander oder sogar in einem einzigen Fragesatz zu stellen.
6. Bei ExpertInneninterviews ist es wichtig die ExpertInnen spüren zu lassen, dass deren Antworten verstanden werden und dass es sich auszahlt den InterviewerInnen Informationen zu geben. Es soll also möglichst viel Kompetenz gezeigt werden.
7. Das Gesagte der InterviewpartnerInnen darf weder negativ noch positiv bewertet werden. Die InterviewpartnerInnen dürfen lediglich darin bestärkt werden, etwas zu erzählen. Bewertet dürfen diese Erzählungen hingegen nicht werden. (vgl. ebd., S.173ff.)

Wie schon erwähnt wurden für die Auswahl der ExpertInnen Kriterien erstellt, nach denen diese ausgesucht wurden. Sobald diese ausgewählt wurden, wurde überlegt, in welcher Form die Interviews stattfinden sollen. Im Idealfall sollten diese face-to-face durchgeführt werden, da dies hauptsächlich Vorteile mit sich bringt. Auch Telefoninterviews bieten Vorteile, wie Zeit- und Kostenersparnisse, jedoch ist bei dieser Art der Befragung weniger Kontrolle gegeben und es kann außerdem zu Informationsverlusten kommen. Überdies ist es bei einem Telefoninterview oft schwer möglich, eine vertrauensvolle Gesprächsatmosphäre herzustellen. (vgl. ebd., S. 153f.) Wurden diese Überlegungen angestellt, so wird mit der Kontaktaufnahme begonnen. Wie schon erwähnt steht hier der Aufbau einer persönlichen Beziehung im Vordergrund. Vor dem Interview gilt es grundsätzlich die ExpertInnen davon zu überzeugen an der Befragung teilzunehmen und zum anderen Vertrauen zu den ExpertInnen aufzubauen. (vgl. ebd., S. 158f.)

Bei dieser Forschungsarbeit wurden die ExpertInnen per E-Mail und per Telefon kontaktiert. Dabei wurde bereits erläutert, um welche Fragestellung es sich bei dieser wissenschaftlichen Arbeit handelt und was das Ziel der Forschung ist. Einige der ExpertInnen wollte auch den Leitfaden im Vorhinein einsehen, so dass dieser teilweise mitgeschickt wurde. Nach der erfolgreichen Kontaktaufnahme, sowie Terminvereinbarung, konnten die ersten Befragungen durchgeführt werden. Obwohl die Interviews idealerweise face-to-face stattfinden sollten, konnten diese

auf Grund bestimmter Umstände nur per Telefon durchgeführt werden. Deshalb konnten auch die Räumlichkeiten für diese Interviews nicht mitbestimmt werden, die womöglich für eine noch bessere Interviewatmosphäre gesorgt hätten. Dennoch konnten die Interviews, die jeweils 20 bis 40 Minuten dauerten, erfolgreich durchgeführt werden.

### 6.3.3 Aufzeichnung und Transkription

Jede Person, die ExpertInneninterviews durchführt, hat eigene Vorlieben, wie das Interview am besten aufgezeichnet werden soll. (vgl. Gläser & Laudel, 2009, S. 157) Bei dieser Forschungsarbeit wurden die Interviews mittels Tonaufzeichnung protokolliert, so dass es zu keinen Informationsverlusten oder -veränderungen kommen konnte. Essenziell war es jedoch, vor der Durchführung der Interviews, zu überprüfen, ob genug Speicherkapazität auf dem Aufnahmegerät vorhanden ist und die Batterien beziehungsweise der Akku aufgeladen sind. Zudem wurde das Aufnahmegerät zuvor getestet, um eine akustische Verständlichkeit gewährleisten zu können. Nachdem alle Interviews durchgeführt wurden, konnte das aufgenommene Material verschriftlicht werden. Dabei galt die Regel: Nicht so viel wie möglich, sondern so viel wie nötig zu transkribieren. Längere Pausen oder spezielle Betonungen, die für die Verständlichkeit notwendig sind, wurden durchaus dokumentiert. Für die Transkription, die mittels des Programms „f4“ durchgeführt wurde, wurden folgende Regeln definiert:

(unv.)	= unverständlich
//	= Überlappung
/	= Abbruch des Satzes
(...)	= Pause
I:	= Interviewerin
B:	= Befragte/r

Tabelle 5: Transkriptions-Regeln – Eigene Darstellung

## 6.4 Erläuterung der Auswertungsmethode

Nach der Aufbereitung der Daten ist der nächste Schritt die Auswertung mittels der qualitativen Inhaltsanalyse nach Mayring. Im Allgemeinen geht es dabei darum, das gesammelte Material zu analysieren (vgl. Mayring, 2015, S. 11). Es handelt sich bei dieser Auswertungsmethode um einen differenzierten und methodisch versierten Umgang mit dem mündlich aufgenommenen Material. (vgl. ebd., S. 26) Die qualitative Inhaltsanalyse ist dahingehend besonders, da die Analyse in einzelne Interpretationsschritte eingeteilt wird, die davor festgelegt werden. (vgl. ebd., S. 61)

Mayring unterscheidet zwischen drei Grundformen der Interpretation: Der Zusammenfassung, Explikation und Strukturierung. Bei dieser Forschungsarbeit wurde die Analysetechnik der Zusammenfassung verwendet. Das Material wurde so reduziert, dass schlussendlich nur mehr die wesentlichen Inhalte übrigblieben. Aus diesem Grund wird hierbei auch von der zusammenfassenden Inhaltsanalyse gesprochen. Wenn bei solchen reduzierenden Textanalyseprozessen nur bestimmte Bestandteile berücksichtigt werden, handelt es sich um die induktive Kategorienbildung. Im Gegensatz dazu gibt es die deduktive Kategorienbildung, bei der das Hauptkategoriensystem im Vorhinein festgelegt wird. Diese wird zumeist bei der Strukturierung angewendet. (vgl. ebd., S. 67f.) Bei der Auswertung des Materials wurde hier die induktive Kategorienbildung angewandt. Laut Mayring (2015) wird dabei folgendermaßen vorgegangen: Sobald anhand den Hauptfragestellungen ermittelt wurde, was zusammengefasst werden soll, wird mit der Bestimmung der Analyseeinheiten begonnen. Darauffolgend werden Kodiereinheiten festgelegt, die paraphrasiert werden sollen. Ist die Paraphrasierung abgeschlossen, so wird mit der Bestimmung des Abstraktionsniveaus und der Generalisierung der Paraphrasen begonnen. Daraufhin wird die erste Reduktion durchgeführt, indem bedeutungsgleiche Paraphrasen gestrichen werden. Falls notwendig kann danach noch eine zweite Reduktion durchgeführt werden. Sind all diese Schritte abgeschlossen, so wird ein neues Kategoriesystem erstellt, das anhand des Ausgangsmaterials kontrolliert wird.

Die eben angeführten Schritte wurden ebenfalls bei der Auswertung der ExpertInneninterviews dieser Arbeit durchgeführt. Dabei wurden zuerst die Hauptfragestellungen festgelegt, um zu bestimmen, was zusammengefasst werden soll:

- Welche Rolle spielen Techniken wie „Artificial Intelligence“ bei der Bekämpfung von Hassreden und in sozialen Netzwerken?
- Wie effizient ist eine Identifikationspflicht bzw. Ausweispflicht, die in Österreich für 2020 vorgesehen ist um gegen Hass im Netz vorzugehen?
- Wieso werden diese Angelegenheiten nicht auf EU-Ebene geregelt?
- Welche Rolle spielt die politische Situation in den einzelnen Ländern, wenn es um die Regulierung von sozialen Netzwerken geht?

Daraufhin wurde bestimmt, sobald das Material zusammengefasst wurde, welche Einheiten für die Analyse notwendig sind. Es wurden also alle Aussagen berücksichtigt, die folgende Punkte ansprachen:

- Das Problem der Hate Speech in sozialen Netzwerken
- Gesetze zur Vermeidung von Hate Speeches
- Einflüsse auf die Regulierung sozialer Netzwerke
- Vorschläge zur Regulierung sozialer Netzwerke
- Regulierungsmethoden der sozialen Netzwerke

Anschließend konnten die relevanten Aussagen aus dem Material bestimmt und paraphrasiert werden. Die Aussagen wurden dabei auf eine knappe und nur auf den Inhalt basierende Form gebracht und daraufhin unter dem Abstraktionsniveau verallgemeinert. Abschließend wurden Paraphrasen gestrichen beziehungsweise gebündelt. Dadurch wurden mehrere Kategorien gebildet.

Bei der Durchführung der Analyse wurde mit einer Tabelle gearbeitet. Diese unterteilt sich in Fall, Seite, Zeile, Zitat, Paraphrase und Generalisierung. Bei dem Fall handelt es sich um Buchstaben, die den jeweiligen Interviews zur Identifizierung zugeteilt wurde.

Fall	Seite	Zeile	Zitat	Paraphrase	Generalisierung	1. Reduktion
LS	1	29-30	Die meisten Fehler, die wir an Hate Speech einstufen oder als Hate Speech wahrnehmen, sind gar keine Hate Speech.	Die meisten Fehler, die wir als Hate Speech einstufen oder wahrnehmen sind gar keine Hate Speech.	Viel was wir als Hate Speech einstufen sind gar keine Hate Speech.	K1: Hate Speech Definition
LS	1	34-37	Es sind Beschimpfungen, es sind ähm unhöfliche Kommentare ähm aggressive Kommentare online und das nennt man eigentlich Online-Harassement. Und wenn Sie in einem Diskussionsforum sind und in diesem Diskussionsforum ähm solche Situationen erleben, dann nennt man das sogar Online-Incivility.	Es sind Beschimpfungen, unhöfliche Kommentare und aggressive Kommentare und das nennt man Online-Harassement. Wenn Sie in einem Diskussionsforum solche Situationen erleben, dann nennt man das Online-Incivility.	Beschimpfungen, unhöfliche und aggressive Kommentare nennt man Online-Harassment. Erlebt man dies in Diskussionsforen nennt man es Online-Incivility.	K2: Hate Speech Abgrenzungen

Abbildung 10: Auswertungs-Tabelle – Eigene Darstellung

#### 6.4.1 Gütekriterien

In der sozialwissenschaftlichen Methodenlehre wird mit zwei Kriterien gearbeitet: der Reliabilität und der Validität. Bei Ersterem geht es um die Stabilität sowie die Genauigkeit der Messung. Die Validität hingegen bezieht sich auf die Gültigkeit der Messung und soll überprüfen, ob auch das gemessen wird, was gemessen werden soll. (vgl. Friedrichs 1973, S. 102, zit.n. Mayring 2015, S. 123) Dadurch, dass diese klassischen Gütekriterien zumeist kritisiert werden, wurden für die qualitative Inhaltsanalyse eigene Kriterien definiert.

Eines davon ist die Verfahrensdokumentation. Dabei geht es hauptsächlich darum, dass Ergebnisse, wissenschaftlich gesehen, wertlos sind, werden diese nicht dokumentiert. Vor allem bei qualitativen Untersuchungen ist dies von großer Bedeutung, da sich dieses Verfahren direkt auf den Untersuchungsgegenstand bezieht. (vgl. ebd., S. 125ff.)

Ein weiteres Kriterium ist die argumentative Interpretationsbeschreibung. Bei der Interpretation ist es besonders wichtig, die Aussagen argumentativ zu begründen, um eine Nachvollziehbarkeit zu gewährleisten. Daher müssen die ForscherInnen theoriegeleitet arbeiten. Es ist hier ebenfalls darauf zu achten, dass Abschnitte, die nicht schlüssig sind, erläutert werden. (vgl. ebd.)

Obwohl die qualitative Forschung eine große Offenheit gewährleistet, ist eine systematische Vorgehensweise essenziell. Vorgegebene Regeln sollten hierbei möglichst eingehalten werden. Aus diesem Grund spricht man hierbei von der Regelgeleittheit. (vgl. ebd.)

Bei qualitativen Forschungen soll vermieden werden, dass sich die interviewten Personen wie „Forschungssubjekte“ fühlen. Deshalb ist es wichtig, dass sich ForscherInnen in die natürliche Lebenswelt der Befragten begeben. Außerdem soll

das Ziel der Interessensübereinstimmung erzielt werden, so dass die größtmögliche Nähe zum Gegenstand erreicht werden kann. Es geht hierbei also um die Gegenstandsangemessenheit. (vgl. ebd.)

Bei dem fünften Kriterium handelt es sich um die kommunikative Validierung. Die Gültigkeit von Ergebnissen der Forschungsarbeit soll gewährleistet sein, daher werden diese den untersuchten Personen vorgelegt. Dies führt zu einer gemeinsamen Diskussion zwischen ForscherInnen und den befragten Personen und dadurch zu der Kontrolle der Ergebnisse. Werden diese vorgelegten Ergebnisse von den befragten Personen bestätigt, so kann die Behauptung aufgestellt werden, dass diese Ergebnisse eine gewisse Gültigkeit aufweisen. (vgl. ebd.)

Das letzte Kriterium, die Triangulation, soll aufzeigen, dass der Versuch aufgestellt werden soll für Fragestellungen verschiedene Lösungswege zu finden und die Ergebnisse daraufhin gegenüberzustellen. Das heißt, dass mehrere unterschiedliche Methoden, Theorien, ForscherInnen oder Datenquellen verwendet werden können. Dies soll die Vor- und Nachteile einzelner Lösungen aufzeigen. (vgl. ebd.)

## 7 Empirische Analyseergebnisse

Durch die Anwendung der qualitativen Inhaltsanalyse konnten die getätigten Aussagen der ExpertInnen in 18 Kategorien eingeordnet werden. Diese wurden wiederum drei Kategoriesystemen zugeordnet. Im nächsten Abschnitt werden diese Kategorien aufgelistet und die entsprechenden Ergebnisse erläutert. Hierbei werden zuvor die jeweiligen Kategorien kurz erklärt um aufzuzeigen, welche Inhalte dabei eingeflossen sind. Daraufhin wird näher auf Aussagen der ExpertInnen eingegangen. Teilweise werden hierbei auch Direktzitate verwendet, wobei die vorgestellten Kürzel der ExpertInnen in der Zitation verwendet werden. Hinter diesem Kürzel wird die Seitenanzahl des jeweiligen Transkripts angegeben um das Direktzitat identifizierbar zu machen.

Die folgende Tabelle soll einen Überblick über die drei Kategoriesysteme und deren untergeordneten Kategorien geben:

Kategoriesystem A	Kategoriesystem B	Kategoriesystem C
<b>Die Problematik der Hate Speech in sozialen Netzwerken</b>	<b>Die derzeitige Situation der Regulierung von sozialen Netzwerken</b>	<b>Maßnahmen für eine effektive Regulierung von sozialen Netzwerken</b>
K1: Hate Speech Definition  K2: Hate Speech Abgrenzung  K3: Ursachen für Hate Speech  K4: Auswirkungen von Hate Speech  K5: Problemwahrnehmung	K6: Regulierung allgemein  K7: Klarnamenpflicht  K8: Netzwerkdurchsetzungsgesetz  K9: Regulierung auf EU-Ebene  K10: Artificial Intelligence als Regulierungsmethode  K11: Richtung der Regulierung	K12: Regulierung auf mehr Ebenen  K13: Voraussetzung & Schwierigkeit der Regulierung  K14: Einfluss auf Regulierung  K15: Notwendigkeit von Bildungsarbeit/Zivilcourage  K16: Notwendigkeit menschlicher Moderation/ExpertInnen  K17: Einsatz von Counter-Speech  K18: Vorschläge & Ideen

Tabelle 6: Kategoriesysteme – Eigene Darstellung

## 7.1 Kategoriesystem A: Die Problematik der Hate Speech in sozialen Netzwerken

Bei diesem Kategoriesystem werden Aussagen bezüglich des Problems mit Online-Hate Speeches behandelt. Dabei geht es hauptsächlich um die Problemwahrnehmung, Ursachen, Auswirkungen, Definition, und Abgrenzungen von Hate Speeches.

### K1: Hate Speech Definition

Bei dieser Kategorie geht es um Aussagen, die sich auf die Definition von Hate Speeches beziehen. Dabei wurde unter anderem die Frage gestellt, wie die jeweiligen ExpertInnen Hate Speech definieren. Die befragten Personen antworteten darauf folgendermaßen: „*Im Bereich Online-Harassment stellt Hate Speech den Bereich dar, wo sich Kommentare gegen Menschen richten auf Grund von Geschlecht, Herkunft und Religion. Es geht um sogenannte Minderheiten. Es sind entweder gruppenbezogene Beschimpfungen oder Hetze.*“ (LS, S. 2) und „*Hate Speech sind alle Äußerungen, die zu Hass anstiften und verhetzen sind oder sich gegen Einzelpersonen oder Gruppen richten auf Grund von Religion, Geschlecht etc. Wir zählen auch Cyber-Mobbing zu Hate Speech*“ (MKT, S. 1f.).

Deutlich wurde dabei, dass es keine einheitliche Definition für diesen Begriff gibt. Häufig fielen Aussagen wie: „*Das Hauptproblem ist die Frage der Definition. Hate Speech ist ein Dachbegriff mit großem Abstraktionsgrad. Viel was wir als Hate Speech einstufen sind gar keine Hate Speech*“ (LS, S.1ff.), „*Bei der Definition haben verschiedene Organisationen unterschiedliche Zugänge*“ (MKT, S. 1) oder „*Es gibt keine eindeutige Definition von Hate Speech.*“ (VF, S. 4f.)

Das Ergebnis dieser Kategorie zeigt also, dass der Begriff der Hate Speech keine eindeutige Definition aufweist und gesellschaftlich unterschiedlich aufgefasst wird. Ersichtlich ist dies vor allem, da es auch in der Forschung anders definiert wird als in der Praxis.

## K2: Hate Speech Abgrenzung

Diese Kategorie bezieht sich ebenfalls auf die Definition von Hate Speech, jedoch geht es hier um die Abgrenzung der Begrifflichkeit. Im Gegensatz zu Hate Speech gibt es auch das Online-Harassment: „*Beschimpfungen, unhöfliche und aggressive Kommentare nennt man Online-Harassment. Erlebt man dies in Diskussionsforen, nennt man es Online-Incivility. Definitorisch ist Online-Incivility auch Hateful Speech, also Hate Speech in Form von Beschimpfungen und eine Bedrohung der Demokratie. Wenn die Kommentare aber empirisch überprüft werden, dann wird es schnell ersichtlich, dass die meisten Fälle von Online Incivility unhöfliche Kommentare sind, kein hate speech oder Bedrohung der Demokratie.*

(LS, S. 1f.) Zudem wird auch zwischen der direkten und der indirekten Hate Speech unterschieden: „*Es gibt den offenen und den indirekten Hate Speech. Der indirekte findet auf einer provokanten, sarkastischen und zynischen Ebene statt.*“ (AW, S. 2)

Es wird hier ersichtlich, dass es ähnliche Phänomene wie Hate Speech gibt, diese jedoch oft falsch eingestuft werden. Zudem wird auch die Hate Speech in zwei Bereiche eingeteilt, die direkte und die indirekte. Deutlich wird hier erneut, dass die Definition von der Hate Speech nicht einheitlich ist.

## K3: Ursachen für Hate Speech

Zum Teil hat man sich bei den Interviews auch mit den verschiedenen Ursachen für die Hate Speech beschäftigt. Von zwei ExpertInnen wurde zum einen erwähnt: „*Die Ursachen können sein, dass der Nutzer ein Troll ist, politische Polarisierung, Lebenssituation des Nutzers oder mangelnde Medienkompetenz.*“ (LS, S. 1f..) Bezuglich dessen wurde jedoch auch betont, dass die Ursachen für Hate Speech genauer untersuchen muss. „*Man fragt sich immer wie bekämpft man das und nicht warum ist das so.*“ (SR, S. 7)

Obwohl hierbei nicht sehr viele Aussagen getätigt wurden, zeigt sich, dass es wichtig wäre sich näher mit den Ursachen auseinanderzusetzen und herausgefunden werden kann, woher der Hass kommt.

#### **K4: Auswirkungen von Hate Speech**

Grundsätzlich wurde hier erwähnt, dass Hate Speeches in der digitalen Welt eine größere Reichweite haben und länger bestehen bleiben als offline. Einigkeit bestand aber vor allem darüber, dass Hate Speeches den öffentlichen Diskurs stark beeinträchtigen. Unter anderem wurde erwähnt: „*Hate Speech hat einen Schlingeffekt auf die Meinungs- und Redefreiheit, weil sich Menschen bedroht fühlen und sich aus dem öffentlichen Diskurs zurückziehen.*“ (FS, S. 2) oder „*Durch Hate-Speech ziehen sich Betroffene aus Online-Debatte zurück.*“ (MKT, S. 3)

Hate Speeches wirken sich also auch auf die Meinungs- und Redefreiheit einzelner Personen aus, da diese sich bedroht fühlen und sich nicht mehr trauen, ihre Meinung zu äußern. Dies hat einen starken Einfluss auf den öffentlichen Diskurs.

#### **K5: Problemwahrnehmung**

Das Problem mit Hate Speech wird grundsätzlich als sehr stark eingestuft. Vor allem durch das Internet ist die Hemmschwelle auf Grund der Anonymität niedriger geworden, so dass Online-Hasskommentare zugenommen haben. Schwierig ist vor allem, dass viele dieser Hassreden nicht erkannt werden. „*Das Problem sind die Hate Speech, die man nicht sofort erkennt.*“ (AW, S. 8) Vor allem die unterschwelligen Hate Speeches sind dabei problematisch. „*Die unterschwellige Hate Speech ist nicht anzeigbar und eine Grauzone und Betroffene können sich schwer dagegen wehren.*“ (AW, S. 2) Auch die organisierten Gruppen werden als sehr problematisch eingestuft: „*Ein Problem sind auch die organisierten Gruppierungen, die gezielt Hate Speech verbreiten und gewisse Inhalte pushen.*“ (VF, S. 7) Obwohl gesellschaftlich noch ein Verständnis dafür fehlt wird die Meinung geteilt, dass „*mehr Aufmerksamkeit auf dem Thema Hate Speech liegt.*“ (VF, S.2)

Zusammenfassend kann hier also gesagt werden, dass die Anonymität des Internets zu einer Senkung der Hemmschwelle geführt hat und sich Hate Speeches dadurch verändert haben. Ein großes Problem ist es, dass

unterschwellige Hassreden eine Grauzone darstellen, bei denen sich Betroffene schwer wehren können. Zudem sind auch die organisierten Gruppierungen ein großes Problem. Gesamtgesellschaftlich muss noch für mehr Bewusstsein gesorgt werden, jedoch ist man der Ansicht, dass heute schon mehr Aufmerksamkeit auf diesem Thema liegt.

## 7.2 Kategoriesystem B: Die derzeitige Situation der Regulierung von sozialen Netzwerken

In diesem Kategoriesystem werden jene Aussagen behandelt, die sich auf aktuelle Regulierungen beziehen. Es geht um die Regulierung allgemein, die Klarnamenpflicht in Österreich, das deutsche Netzwerkdurchsetzungsgesetz, die Regulierung auf EU-Ebene, Artificial Intelligence und um die Richtung der bisherigen Regulierungen.

### K6: Regulierung allgemein

In dieser Kategorie werden allgemeine Aussagen zu der Regulierung von sozialen Netzwerken bezüglich Hate Speech dargestellt. Grundsätzlich ist man der Ansicht, dass noch zu wenig gelöscht wird: „*Bei der Bekämpfung gegen Hate Speech ist bei Facebook Twitter und Co. noch Luft nach oben. Es wird einfach noch zu wenig gelöscht und nicht sensibel genug mit der Thematik umgegangen.*“ (MKT, S. 3f.) Die befragten Personen merken hier an, dass es zu wenige und zu unausbildete MitarbeiterInnen bei den sozialen Netzwerken gibt, die sich darum kümmern. „*Das Problem ist, dass es viele Meldungen sind und zu wenige Mitarbeiter die manuell nach ihren eigenen Gutdünken entscheiden was Hate Speech ist und was nicht.*“ (AW, S. 3f.) „*Es braucht ausgebildete und trainierte Personen die bei den Plattformen dafür zuständig sind.*“ (MKT, S. 4)

Als Ergebnis kann hier also gesagt werden, dass die momentane Vorgehensweise noch nicht zufriedenstellend ist, da noch zu wenig gelöscht wird. Es wird hier an die sozialen Netzwerke appelliert, mehr ausgebildete Personen einzustellen, die sich um die Löschung von Hassreden kümmern.

### **K7: Klarnamenpflicht**

Bei dieser Kategorie wurden alle Aussagen der befragten Personen zusammengefasst, bei denen es sich um die geplante Klarnamenpflicht in Österreich handelte. Grundsätzlich wird hier folgende Meinung geteilt: „*Die Auskunftspflicht/Klarnamenpflicht in Österreich ist nicht zielführend gegen Hass im Netz. Diese hilft weder den respektvollen Umgang im Internet zu fördern noch erleichtert das die Betroffenen Rechtsansprüche zu verfolgen.*“ (MKT, S. 6) „*Die Auskunftspflicht ist reiner Aktionismus. In Südkorea hat das zum Beispiel gar nicht funktioniert. Man hat das Recht auf Anonymität im Netz. Manche Leute müssen auch anonym bleiben, wenn sie etwas veröffentlichen.*“ (AW, S. 4)

Dabei wurde häufig erwähnt, dass hierbei von einer falschen Prämisse ausgegangen wird, da die meisten HassposterInnen unter ihrem Klarnamen posten: „*Die Auskunftspflicht geht am Kern des Problems vorbei. Viele Hater posten unter Klarnamen.*“ (VF, S. 3) „*Das österreichische Vermummungsverbot bringt wenig, da die meisten unter Klarnamen posten. Es ist ein Schlingeffekt auf die Meinungsfreiheit.*“ (FS, S. 6)

Die meisten ExpertInnen sind sich hier also einig, dass die Klarnamenpflicht in Österreich am Kern des Problems vorbei geht, da die meisten HassrednerInnen sowieso schon unter ihrem Klarnamen posten. Zudem haben auch zwei der ExpertInnen (AW, VF) erwähnt, dass so ein ähnliches Gesetz auch in Südkorea beschlossen wurde und nicht sehr gut funktionierte, da es zu Daten-Leaks führte. Aus diesem Grund wird dieses Gesetz als nicht sehr zielführend gesehen.

### **K8: Netzwerkdurchsetzungsgesetz**

Bei dieser Kategorie geht es um die ExpertInnenaussagen bezüglich des deutschen Netzwerkdurchsetzungsgesetzes. Grundsätzlich konnten sich alle ExpertInnen zu diesem Gesetz äußern. Negativ an diesem Gesetz ist zum einen: „*Bei dem NetzDG entscheiden nicht Richter, sondern die Plattformen selber was illegal ist und was nicht. Es ist nicht angemessen, Entscheidungen über Hate Speeches in den privaten Bericht zu delegieren.*“ (FS, S. 7f.) Zwei weitere Expertinnen waren ähnlicher Ansicht: „*Negativ ist, dass die Rechtsdurchsetzung*

*privatisiert wurde“.* (VF, S. 3) „Das Problem bei dem NetzDG ist, dass irgendjemand bei Facebook und Co. entscheiden muss, was Volksverhetzung ist und was nicht. Da eröffnet sich die Möglichkeit der Einschränkung der Meinungsfreiheit.“ (LS, S. 6)

Zum anderen herrscht im Allgemeinen Einigkeit darüber, dass „Das Problem des NetzDG ist das Overblocking, dass eine Gefahr für die Meinungsfreiheit darstellt.“ (FS, S. 7) „Durch das NetzDG kann es sein, dass die Meinungsfreiheit eingeschränkt wird, weil Plattformunternehmen eher mehr löschen.“ (MKT, S. 7)

Einige ExpertInnen äußersten sich jedoch auch positiv zu diesem Gesetz. Eine Expertin erwähnte beispielsweise: „Das Netzwerkdurchsetzungsgesetz ist ein Indikator dafür, dass die Politik noch Regulierungsmacht gegenüber sozialen Netzwerken hat.“ (LS, S. 7) Ebenfalls positive empfindet man, dass dadurch zur Aufmerksamkeit geschaffen wurde: „NetzDG ist positiv, da es zur Sensibilisierung beigetragen hat und Plattformen mehr Personal eingestellt haben“. (VF, S. 3)

Zwei der ExpertInnen haben jedoch auch erwähnt, dass es ein Gesetz ist, dass nicht unbedingt gebraucht worden wäre, da die Rechtsgrundlage schon vorhanden war. „Das NetzDG ist ein wichtiges Gesetz, dass es nicht unbedingt gebraucht hätte, weil die Rechtsgrundlage war da. Es geht hier um die Durchsetzung solcher Sachen.“ (SR, S. 5)

Als Ergebnis kann hier festgehalten werden, dass das Netzwerkdurchsetzungsgesetz zwar ein Ansatz ist, jedoch zwei Hauptprobleme aufweist. Zum einen nämlich das Problem, dass die Rechtsdurchsetzung privatisiert wurde und zum anderen, dass die Gefahr des Overblocking herrscht, das dazu führt, dass die Meinungsfreiheit eingeschränkt wird. Jedoch sehen es zwei ExpertInnen auch positiv, da durch dieses Gesetz Aufmerksamkeit zu diesem Thema geschaffen wurde sowie gezeigt wurde, dass die Politik sehr wohl eine Regulierungsmacht gegenüber den sozialen Netzwerken hat.

## **K9: Regulierung auf EU-Ebene**

Diese Kategorie bezieht sich auf die Frage, warum dieses Problem der Hate Speech nicht auf EU-Ebene reguliert wird. Hierbei waren sich die ExpertInnen nicht einig. Eine Expertin erklärte beispielsweise: „*Die Regulierung von Hate Speech innerhalb der EU hat auch mit historischen Erfahrungen und Rechtssystemen in den unterschiedlichen Ländern zu tun.*“ (LS, S. 7) Das heißt, dass eine Regulierung auf EU-Ebene insofern schwer ist, da jedes Land ein anderes Rechtssystem hat. Das Verhetzungsverbot hat zum Beispiel viel mit der Geschichte des Holocausts zu tun. Aus diesem Grund ist dieses Gesetz in Deutschland und Österreich sehr eng gefasst. Daher wird hier die Meinung geteilt, dass eine EU-Regulierung schwierig sei.

Eine weitere Expertin deutet an, dass „*eine EU-Regelung nicht unbedingt notwendig ist, sondern es muss ein Bekenntnis der nationalen Politik her wo dann auch Taten folgen.*“ (MKT, S. 8) Hier ist man ebenfalls gegen eine EU-Regelung, da eher an die nationale Politik appelliert wird.

Zwei weitere Experten sind sich nicht sicher, ob eine EU-Regelung möglich ist. Dies hängt stark von den verschiedenen Interessen der Staaten ab. „*Bei EU-Regulierung muss man sich anschauen wer hat Interesse an Aufstellung eines Gesetzes und wer nicht.*“ (AW, S. 7) „*Eine EU-weite Regelung hängt mit den Nationalstaatsinteressen zusammen.*“ (SR, S. 6)

Allgemein wird einer EU-Regulierung also eher skeptisch gegenüber gestanden beziehungsweise hält dies nicht für notwendig. Lediglich eine Expertin (VF) fände eine Regelung auf EU-Ebene sinnvoll.

## **K10: Artificial Intelligence als Regulierungsmethode**

Diese Kategorie gibt einen Überblick darüber, was die ExpertInnen von Artificial Intelligence als Regulierungsmethode halten. Hierbei sind die meisten der befragten Personen der Meinung, dass die künstliche Intelligenz für die Regulierung nicht sehr geeignet ist. „*Machine-Learning Prozesse eignen sich nicht direkt für die Regulierung von Hate Speech, da sie den Kontext nicht*

erfassen können.“ (LS, 2f.) „Algorithmen verstehen nicht wenn jemand einen Hate Speech zitiert oder diesen umdreht und als MEME veröffentlicht. Das ist das Problem mit Künstlicher Intelligenz.“ (AW, S. 3) „Eine sehr gut lernen Software, wird den Menschen in Bezug auf Content Moderation noch lange nicht komplett ersetzen können.“ (SR, S. 3) „Muss noch viel passieren, dass soziale Netzwerke Artificial Intelligence sinnvoll einsetzen können.“ (MKT, S. 4)

Ein Experte erwähnte jedoch auch, dass diese Art der Regulierung für soziale Netzwerke günstiger sei, als menschliche Moderation. „Soziale Netzwerke setzen künstliche Intelligenz ein, da es kostengünstig und standardisiert ist.“ (FS, S. 3f.) Dieser ist jedoch auch der Meinung, dass die Erkennung von Hate Speech nicht voll automatisierbar ist.

Von dem Einsatz der künstlichen Intelligenz der sozialen Netzwerke zur Erkennung von Hate Speeches halten die ExpertInnen nicht sehr viel. Es herrscht Einigkeit darüber, dass eine solche Software noch viel lernen muss, bis diese die menschliche Moderation komplett ersetzen kann.

### **K11: Richtung der Regulierung**

Bei dieser Kategorie geht es um die Frage, in welche Richtung sich die Regulierung momentan bewegt. Es haben sich dabei zwei ExpertInnen dazu geäußert. Zum einen wurde erwähnt: „Die Tendenz ist es staatlich einzugreifen. Also das NetzDG geht in Richtung staatliche Regulierung. Also es geht schon eher in die Richtung staatliche Regulierung.“ (FS, S. 8f.) Eine weitere Expertin äußerte sich bezüglich dessen folgendermaßen: „Ich glaube, dass Selbstregulierung alleine nicht ausreicht. Ich glaube man kann auch noch positive Ansätze setzen für Selbstregulierungsmechanismen oder für sowas wie Code of Conduct oder ethische Kodizes auferlegen.“ (VF, S. 5)

Obwohl sich hierzu nur zwei ExpertInnen geäußert haben herrschte Einigkeit, dass eine Selbstregulierung nicht ausreicht. Ähnlich wie das Netzwerkdurchsetzungsgesetz wird die Meinung geteilt, dass es auch zukünftig eher in die Richtung der staatlichen Regulierung gehen wird.

## 7.3 Kategoriesystem C: Maßnahmen für eine effektive Regulierung von sozialen Netzwerken

Nach dem Erfassen von Informationen über die Hate Speech allgemein sowie über die aktuellen Regulierungsversuche, geht es hier nun um die Darstellung von Aussagen, die sich auf mögliche Maßnahmen für eine effektive Regulierung beziehen. Es wurden hierbei Daten gesammelt bezüglich Regulierungsebenen, Voraussetzungen und Schwierigkeiten, Einflüsse auf die Regulierung, die Notwendigkeit von Bildungsarbeit und menschlicher Moderation, der Einsatz von Counter-Speech sowie Vorschläge und Ideen für eine effektive Regulierung.

### K12: Regulierung auf mehreren Ebenen

Bei der Frage nach Maßnahmen für eine effektiver Regulierung erwähnten vier der sechs ExpertInnen, dass eine Regulierung auf zwei Ebenen stattfinden muss. „*Online-Hate Speeches sind ein soziotechnischer Prozess. Die Regulierung kann nicht nur auf Ebene des Sprechers laufen, sondern es müssen die technischen Prozesse ebenfalls berücksichtigt werden.*“ (LS, S. 4) „*Bei Regulierung muss auf Struktur von sozialen Netzwerken eingegangen werden.*“ (FS, S. 3) „*Man muss auf die Techniken von sozialen Netzwerken eingehen.*“ (MKT, S. 10) „*Man sollte sich bei der Regulierung auch anschauen wie soziale Netzwerke funktionieren und wie wird dort Meinung gebildet wird.*“ (VF, S. 6)

Als Beispiel wurden hier von einem Experten die Algorithmen genannt. „*Algorithmen spielen bei Hate Speeches zentral rein, da diese emotionalisierend sind und so meist viel Aufmerksamkeit bekommen.*“ (FS, S. 3)

Das Ergebnis in dieser Kategorie zeigt also, dass eine Regulierung nicht nur auf Ebene der NutzerInnen, sondern auch auf Ebene der sozialen Netzwerke stattfinden muss. Es wird erwähnt, dass bei der Regulierung auf die Struktur von sozialen Netzwerken eingegangen werden muss.

### **K13: Voraussetzung & Schwierigkeit der Regulierung**

In dieser Kategorie werden Aussagen zusammengefasst, die sich auf die Voraussetzungen und Schwierigkeiten der Regulierung beziehen.

Hierbei wurden von den ExpertInnen viele verschiedene Aussagen getätigt. Unter anderem wurde erwähnt, dass für die Regulierung ein guter Einblick in die Konfliktdynamik gegeben sein muss. „*Man muss die Regulierung auf verschiedenen Ebenen einsetzen. Eine davon sind User und eine davon Plattformen. Eine ersetzt nicht die andere und um das zu schaffen braucht man einen guten Einblick in die Konfliktdynamik.*“ (LS, S. 4) Dieselbe Expertin erwähnte ebenfalls: „*Globalagierende Konzerne kann man nicht auf nationaler Ebene regulieren. Regulierung von global-agierenden Online-Netzwerk-Plattformen auf nationaler Ebene wird auf klare Grenzen stoßen.*“ (LS, S. 8) Sie warf außerdem in den Raum, ob Hate Speeches überhaupt reguliert werden sollen, „*denn Staaten können Hate Speech Gesetze instrumentalisieren*“. (LS, S. 8)

Es wurde hier ebenfalls erneut erwähnt, dass es Voraussetzung sei auf die Strukturen von sozialen Netzwerken einzugehen. „*Bei solchen Regulierungen sollte man auf die Struktur von sozialen Netzwerken eingehen oder Regelungen finden, die diesen Netzstrukturen und Dynamiken besser entsprechen.*“ (FS, S. 3) Dieselbe Person ergänzte: „*Es gibt keine One-Size-Fit-All-Lösung. Es gibt nicht die eine Lösung, mit der man alle Probleme löst. Es gibt nur verschiedene Ansätze, mit denen man einen bestimmten Nutzer erzielt und da muss man schauen, wie groß der Schaden ist.*“ (FS, S. 7)

Zwei weitere ExpertInnen sind der Meinung, dass es wichtig sei, dass bei der Regulierung mehrere Komponenten zusammenspielen müssen. „*Bei der Regulierung müssen Plattformen, Beratungsstellen und die Politik zusammenspielen.*“ (MKT, S. 9) „*Man muss einen passenden Weg finden, wo alle Komponente abgedeckt und eingebaut sind.*“ (AW, S. 7f.)

Innerhalb dieser Kategorie wurden von vier ExpertInnen ganz unterschiedliche Aussagen darüber getätigt, auf was bei der Regulierung geachtet werden muss und was die Schwierigkeit dabei ist. Zum einen wurde erwähnt, dass es schwierig

ist, ein globales Plattformunternehmen auf nationaler Ebene zu regulieren. Zudem wurde erwähnt, dass eine Balance zwischen UserInnen und der Plattform und deren Konfliktodynamik gefunden werden muss. Wichtig bei der Regulierung ist es ebenfalls, dass mehrere AkteurlInnen, also Plattformen, Politik und Meldestellen zusammenarbeiten.

#### **K14: Einfluss auf Regulierung**

Diese Kategorie befasst sich mit der Frage, welche Einflüsse bei der Regulierung von sozialen Netzwerken bezüglich Hate Speech einwirken. Hierbei wurde vor allem erwähnt, dass die Regierungen in den verschiedenen EU-Staaten einen großen Einfluss darauf haben. „*Ich glaube schon, dass die Politik in einem Land Einfluss auf die Regulierung hat, zumindest auf legislativer Ebene.*“ (FS, S. 9f.) „*Ich glaube ganz gewiss, dass die politische Situation in einem Land Einfluss auf diese Regulierung hat.*“ (AW, S. 7) „*Regulierungen haben immer mit politischer Situation in einem Land zu tun.*“ (SR, S. 6)

Zwei ExpertInnen waren hier sogar der Ansicht, dass rechtsextreme oder konservative Parteien eher gegen so eine Regulierung sind. „*Ich glaube, dass es möglich ist, dass konservative Regierungen und auch rechtspopulistische Regierungen weniger Interesse an einer Regulierung haben.*“ (AW, S. 7) „*Man kann gut beobachten, dass rechte Parteien oder rechtsextreme Parteien oder Parteien, die eher in Richtung rechtsextrem tendieren, eigentlich gegen eine Regulierung von Hate Speech sind. Wenn diese Parteien auch einen politischen Einfluss haben dann hat es einen Einfluss.*“ (VF, S. 5f.)

Daraus lässt sich schließen, dass der Großteil der Befragten der Meinung ist, dass die politische Situation in einem Land einen Einfluss auf die Regulierung von sozialen Netzwerken bezüglich Hate Speeches hat. Vor allem bei konservativen oder rechtsextremen Parteien besteht die Möglichkeit, dass diese eher gegen so eine Regulierung sind.

## K15: Notwendigkeit von Bildungsarbeit/Zivilcourage

Diese Kategorie gibt einen Überblick über die Notwendigkeit von Bildungsarbeit und Zivilcourage. Alle ExpertInnen waren sich dabei einig, dass es notwendig ist Bildungsarbeit zu leisten und dabei die Medienkompetenz zu erhöhen, sowie die Zivilcourage zu steigern. „Das Problem löst man nicht nur mit gesetzlichen Regulierungen. Es geht auch um die Erhöhung der Medienkompetenz, vor allem bei älteren Leuten.“ (LS, S. 2) „Es braucht auch Zivilcourage, dass Betroffene nicht alleine gelassen werden und dass Menschen eingreifen.“ (MKT, S. 5) „Ich glaube es braucht eine Förderung von Beratungsstellen, es braucht Förderung von Monitoring-Mechanismen, es braucht Medienbildung, Medienkompetenz.“ (VF, S. 5) „Es ist definitiv eine Frage der Medienkompetenz. Ich glaube, dass das eine der Hauptgründe ist, warum es so häufig auftritt.“ (SR, S. 2) „Es ist Zivilcourage und Gespür für Menschenwürde, wenn man mit Counter-Speech Hass verdünnt.“ (FS, 5)

Hierbei wurde auch von zwei ExpertInnen ergänzt, dass es auch TäterInnenarbeit benötigt. „TäterInnen fehlt oft das Bewusstsein, dass diese etwas falsch machen.“ (MKT, S. 4) „Es braucht Täterarbeit, weil ich glaube, dass es vielen nicht bewusst ist, dass es strafrechtliche Konsequenzen heben kann, wenn sie Hate Speech online tätigen.“ (VF, S. 2)

Zur Frage, wie man die Medienkompetenz am besten steigert, wurde erklärt: „Eine der Wege wäre, wenn man Medienaufklärung mehr in Schulen einbeziehen würde. Es wäre ein wichtiger Schritt.“ (SR, S. 2) „Es gibt schon das ZARA-Training, wo an viel mit Schulen gearbeitet wird. Das ist ein wichtiger Punkt, weil Kinder und Jugendliche wachsen mit Hass im Netz auf und sollten wissen wie sie digitale Zivilcourage ausüben können.“ (MKT, S. 10) Es wurde jedoch von einer Expertin erwähnt, dass dafür eine finanzielle Unterstützung notwendig ist. „Es muss noch viel Bildungsarbeit, Bewusstseinsarbeit und Sensibilisierungsarbeit passieren und dafür muss es Geld geben.“ (MKT, S. 10)

Es wird hier sehr deutlich, dass dies auch eine Frage der Medienkompetenz ist sowie der Zivilcourage. Aus diesem Grund wäre eine Maßnahme gegen Hate

Speech vorzugehen indem mehr Bildungsarbeit- und Bewusstseinsarbeit an Schulen zu geleistet wird.

### **K16: Notwendigkeit menschlicher Moderation/ExpertInnen**

Bei dieser Kategorie werden alle Aussagen zusammengefasst, die sich mit der Notwendigkeit von menschlicher Moderation befassen. Hierbei herrschte mehrheitlich Einigkeit, dass es die menschliche Moderation beziehungsweise ExpertInnen benötigt um gegen Hate Speech vorzugehen. „*Sinnvoll wäre es ExpertInnen in die Unternehmen zu setzen, die dann wirklich dagegen vorgehen.*“ (MKT, S. 7) „*Zusätzlich zu der Künstlichen Intelligenz mehr Leute einstellen wäre gut, denn ein Mensch kann im Idealfall sarkastische Untertöne erkennen.*“ (AW, S. 4) „*Es braucht ein Expertenteam, dass sich auskennt und wo die Möglichkeit da ist, das dann zu verarbeiten.*“ (SR, S. 3) „*Es braucht Menschen vor Ort, die die Kultur kennen, die Sprache verstehen und Satire richtig einordnen können und dann Entscheidungen treffen.*“ (FS, S. 5)

Zusätzlich wurde von einem Experten erwähnt, dass menschliche Moderation zwar notwendig wäre, den sozialen Netzwerken jedoch auch teurer kommen würde als die Anwendung von AI. „*Bei der Erkennung von Hate Speech braucht es viel menschliche Moderation um diese Kontexte gut zu verstehen. Es ist wichtig da Leute einzusetzen, die den Kontext und die Sprache verstehen. Da wird es aber dann im Vergleich zu den Maschinen relativ teuer.*“ (FS, S. 4) Aus diesem Grund wäre der Vorschlag, eine Kombination aus menschlicher Moderation und AI. „*Unter dem Strich würde man eine Mischung von künstlicher Intelligenz und menschlicher Moderation zur Erkennung von Hate Speech anwenden.*“ (FS, S. 4f.)

Zusammenfassend kann hier also gesagt werden, dass es menschliche Moderation bedarf. Es sollen mehr ExpertInnen in den Plattformunternehmen eingesetzt werden, die den Kontext eines Posts richtig einordnen können.

## K17: Einsatz von Counter-Speech

Diese Kategorie beinhaltet die Aussagen bezüglich des Einsatzes von Counter-Speech. Hierzu haben sich zwei ExpertInnen geäußert. Eine Person betonte: „*Es ist wichtig den Hass zu verdünnen. Dass die Zivilgesellschaft, wenn die etwas sehen dagegenreden und den Hass verdünnen. Also dass man Hass mit positiven Botschaften ausdünnnt, so dass dieser untergeht.*“ (FS, S. 5) Eine weiter Expertin erwähnte, dass ihre Organisation bereits mit Counter-Speech arbeitet „*Wir wenden auch die Gegenreden in verschiedenen Formen an um auch ein bisschen ein Vorbild zu sein und zu zeigen was für Möglichkeiten man hat und auch weil es wichtig ist, wenn man ein Social Media Forum betreibt.*“ (MKT, S. 3)

Um gegen Hassreden vorzugehen wäre die Anwendung von Counter-Speech also eine Möglichkeit. Außerdem wird eine Vorbildfunktion gegenüber anderen UserInnen übernommen, so dass diese die Gegenrede ebenfalls anwenden.

## K18: Vorschläge & Ideen

Bei dieser Kategorie werden Aussagen der ExpertInnen aufgelistet, die Vorschläge und Ideen für eine effektive Regulierung beinhalten. Es wurde unter anderem erwähnt, dass beispielsweise Gremien errichtet werden sollen: „*Es braucht eine zivilgesellschaftliche Zwischenlösung, wo man beispielsweise Gremien installiert, die bunt besetzt sind und dann solche Empfehlungen abgeben, wie man mit solchen Inhalten umgehen sollte, was man löschen, sperren, etc. soll.*“ (FS, S. 9) „*Vielleicht wäre es gut eine unabhängige rechtliche Stelle einzusetzen, die das noch einmal prüfen kann.*“ (VF, S.4) „*Es sollte eine unabhängige Stelle geben, die diese Kommentare oder Postings einschätzen kann.*“ (MKT, S. 8) Zudem wurde eingebbracht, dass eine weltweite Regulierung am sinnvollsten wäre. „*Tatsächlich wäre die Lösung ein weltweiter Konsens, weil das Internet hört ja auch nicht irgendwie an europäischen Grenzen auf.*“ (SR, S. 6)

Zudem wurde erwähnt, dass es wichtig wäre an der richtigen Stelle anzusetzen. „*Es gibt oft das Problem, dass die Betroffenen alleine gelassen werden und sich nicht leisten können dagegen vorzugehen, nicht die Zeit und Energie haben. Das sind die Sachen, wo man ansetzen muss.*“ (MKT, S. 6)

Als Ergebnis ist zu sagen, dass es von den ExpertInnen unterschiedliche Ideen gibt, wie soziale Netzwerke regulieren werden können um gegen Hate Speeches vorzugehen. Zwei ExpertInnen sind jedoch der Meinung, dass eine unabhängige Stelle beziehungsweise ein Gremium, die Hate Speeches prüft, sinnvoll wäre.

## 7.4 Interpretation der Ergebnisse

Im Rahmen der Erhebungsmethode wurden verschiedene Themen angesprochen, die sich mit Hate Speeches sowie deren Regulierung befassen. Bei der **Problematik mit Hate Speeches** wurde vorrangig das Problem mit der Definition angesprochen. Mehrmals wurde erwähnt, dass es keine eindeutige Definition für diesen Begriff gibt, was sich in der Hinsicht als problematisch erweist, da grundsätzlich nichts reguliert werden kann, wenn nicht gewusst wird was reguliert werden soll. Zudem wurde hier erläutert, dass ein weiteres Problem, das sei, dass es unterschwellige Hate Speeches gibt, die nicht sofort als solche identifiziert werden. Diese bilden eine Grauzone und können für Betroffene eine große Belastung sein, da diese sich dagegen nur schwer wehren können. Eine weitere Problematik der Hate Speech ist, dass dadurch der öffentliche Diskurs beeinträchtigt wird. Dadurch, dass sich viele UserInnen dadurch bedroht fühlen, haben sie Angst sich weiterhin zu äußern und ziehen sich deshalb aus Online-Debatten zurück. Im Allgemeinen kann hier also gesagt werden, dass diese Problematik der Hate Speech als sehr stark wahrgenommen wird.

Derzeit wird anhand nationaler Gesetze versucht, gegen Hate Speeches in sozialen Netzwerken vorzugehen. Bei **der derzeitigen Situation der Regulierung von sozialen Netzwerken** gibt es definitiv noch Luft nach oben, da momentan noch zu wenig gelöscht wird und nicht sensibel genug mit diesem Thema umgegangen wird. Obwohl Österreich eine Klarnamenpflicht zur Regulierung einführen möchte, herrscht hier Einigkeit, dass dies zu keiner Lösung des Problems führen wird. Dies ist hauptsächlich deswegen der Fall, da die meisten HassposterInnen heute unter Klarnamen posten. Außerdem sollten InternetnutzerInnen das Recht auf Anonymität haben, da manche UserInnen sogar anonym bleiben müssen, wenn diese etwas veröffentlichen wollen

beziehungsweise müssen. Zudem wird auch deswegen nicht sehr viel von diesem Gesetz gehalten, da solches bereits in Südkorea eingeführt wurde und zu großen Daten-Leaks führte. Ein weitaus besserer Ansatz ist das deutsche Netzwerkdurchsetzungsgesetz. Obwohl dies auf mehr Zuspruch stieß hat man auch hier Bedenken, dass ein Overblocking zu einer Einschränkung der Meinungsfreiheit führt. Des Weiteren wird bei diesem Gesetz als kritisch betrachtet, dass nicht die RichterInnen, sondern die Plattformen selber darüber entscheiden, was illegal ist und was nicht. Eine Regulierung auf EU-Ebene wird von den ExpertInnen teilweise als sehr sinnvoll und teilweise als nicht notwendig empfunden. Obwohl noch viel Kritik an diesen Gesetzen ausgeübt wird ist man der Meinung, dass es in Richtung staatliche Regulierung gehen wird.

Von dem Einsatz von AI zur Erkennung von Hate Speeches halten die befragten Personen nicht sehr viel, da diese Technik noch nicht so ausgereift ist, dass sie den Kontext eines Postings richtig erfassen könnte. Zwar ist dies für die sozialen Netzwerke günstiger, als wenn mehr Personal eingestellt wird jedoch wird die künstliche Intelligenz die menschliche Moderation noch lange nicht komplett ersetzen können.

Mögliche **Maßnahmen für eine effektive Regulierung von sozialen Netzwerken** wäre es, wenn mehr Bildungsarbeit in Schulen geleistet werden würde um so die Medienkompetenz sowie die Zivilcourage bei Kindern und Jugendlichen zu steigern, so dass möglicherweise die Anzahl von Hate Speeches in den sozialen Netzwerken gesenkt werden. Des Weiteren sollte die Counter Speech mehr zum Einsatz kommen, so dass der Hass im Netz eventuell verdünnt werden kann. Eine weitere wichtige Maßnahme wäre es, wenn mehr ExpertInnen in den Plattformunternehmen sitzen würden, die Hate Speech Meldungen überprüfen. Es benötigt die menschliche Moderation um den Kontext richtig einordnen zu können. Essenziell bei der Regulierung ist ebenfalls, dass nicht nur auf der Seite der UserInnen reguliert wird, sondern vielmehr auch auf die Architektur von Plattformen eingegangen wird. Es muss verstanden werden, wie soziale Netzwerke funktionieren, so dass dort angesetzt und reguliert werden kann. Die Schwierigkeiten bei so einer Regulierung sind jedoch, dass ein passender Weg gefunden werden muss, wo alle Komponenten abgedeckt sind. Das heißt es kann

nicht nur eine Partei regulieren, sondern es müssen alle Parteien, also Politik, Plattformunternehmen sowie Beratungsstellen zusammenhelfen. Des Weiteren wird die Ansicht geteilt, dass man bei der Regulierung eines global-agierenden Konzerns auf klare Grenzen stoßen wird. Grundsätzlich ist jedoch zu erwähnen, dass es keine One-Size-Fit-All-Lösung geben wird. Es ist hierbei wichtig, verschiedene Ansätze abzuwägen und so analysieren wie groß der Nutzen und wie groß der Schaden ist. Dies stellt am Ende eine der größten Herausforderungen dar.

## 7.5 Beantwortung der Forschungsfragen

Nachdem nun die Ergebnisse der qualitativen Analyse aufgelistet und beschrieben wurden, können damit nun die vier Forschungsfragen beantwortet werden. Am Ende soll damit auch die Leitfrage beantwortet werden können.

### 7.5.1 Forschungsfrage 1

#### **FF1: Welche Rolle spielen Techniken wie „Artificial Intelligence“ bei der Bekämpfung von Hassreden und in sozialen Netzwerken?**

Grundsätzlich ist die Methode der Artificial Intelligence zur Erkennung von Hate Speech eine kostengünstige Lösung für die sozialen Netzwerke. Die ExpertInnen sind sich hier jedoch einig, dass AI zur Erkennung von Hate Speeches noch nicht so weit ist, da diese nicht jeden Kontext eines Postings versteht. Dies führt dazu, dass diese entweder zu viele Inhalte löschen, was zu einer Einschränkung der Meinungsfreiheit führt, oder zu wenig. Die niedrige Erfolgsquote zeigt, dass die Erkennung von Hate Speech noch nicht voll automatisierbar ist. Bis sich hierbei vollkommen auf AI verlassen werden kann, muss noch viel passieren. Aus diesem Grund ist eine menschliche Moderation dringend notwendig. Es müssen für die Erkennung von Hate Speech Personen eingesetzt werden, die den Kontext verstehen. Vor allem braucht es auch Personen vor Ort, die die Sprache verstehen, die Kultur kennen und Satire richtig einordnen können. Es wird gefordert, dass die sozialen Netzwerke mehr ExpertInnen in diesem Bereich einstellen, die sich mit der Thematik auskennen und einordnen können, was Hate Speeches sind und was nicht. Dafür muss eine gewisse Vision und Strategie da

sein um zu wissen, was diese MitarbeiterInnen brauchen und wie diese eingesetzt werden sollen. Grundsätzlich lässt sich hier also zusammenfassend sagen, dass Artificial Intelligence momentan noch nicht so weit ist um Hate Speeches ohne jegliche Hilfe identifizieren zu können. Ideal wäre deshalb eine Mischung aus AI und menschlicher Moderation.

### 7.5.2 Forschungsfrage 2

#### **FF 2: Wie effizient ist eine Identifikationspflicht bzw. Ausweispflicht, die in Österreich für 2020 vorgesehen ist um gegen Hass im Netz vorzugehen?**

Bei der Frage nach der geplanten Klarnamenpflicht in Österreich herrscht Einigkeit darüber, dass dies kein angemessenes Gesetz darstellt, das zur Bekämpfung von Hate Speech in sozialen Netzwerken beiträgt. Das Speichern von Daten der NutzerInnen führt vielmehr dazu, dass diese sich nicht mehr trauen, sich zu bestimmten Themen zu äußern. Des Weiteren soll das Recht auf Anonymität im Netz bestehen, da gewissen NutzerInnen bei der Veröffentlichung gewisser Inhalte unbekannt bleiben müssen. Grundsätzlich geht dieses Gesetz jedoch von einer völlig falschen Annahme aus. Es wird hier behauptet, dass HassposterInnen so schneller zur Rechenschaft gezogen werden können. Jedoch ist es so, dass die meisten HassrednerInnen heute sowieso unter ihren Klarnamen posten. Diese Auskunftspflicht beziehungsweise Klarnamenpflicht geht deswegen am Kern des Problems vorbei und ist nur wenig zielführend.

Die befragten ExpertInnen sind der Ansicht, dass das Netzwerkdurchsetzungsgesetz in Deutschland ein besserer Ansatz sei. Hierbei sollen offensichtlich rechtswidrige Inhalte innerhalb von 24 Stunden gelöscht werden. Obwohl dies auf mehr Zuspruch stößt, hat man auch hier Angst vor einem sogenannten Overblocking, das ebenfalls zu einer Einschränkung der Meinungsfreiheit führen kann. Des Weiteren ist es fragwürdig, ob es sinnvoll ist, die Rechtsdurchsetzung zu privatisieren. Denn wie schon erwähnt entscheiden nicht RichterInnen, sondern eben die sozialen Netzwerke, was als Hate Speech eingestuft wird und was nicht. Zumal hier auch ein grundsätzliches Problem ist, dass es keine einheitliche Definition von Hate Speech gibt. Einmal abgesehen davon ist es aber auch notwendig, ExpertInnen in den sozialen Netzwerken einzusetzen, die Hate Speeches richtig einstufen können.

Bei diesen Regulierungen wurde bislang nur auf einer Ebene angesetzt. Laut der befragten ExpertInnen wäre es hingegen sinnvoller nicht nur auf Ebene der NutzerInnen zu regulieren, sondern auch auf die Struktur von sozialen Netzwerken einzugehen. Hierbei wurde erwähnt, dass vor der Regulierung verstanden werden muss wie die Logiken von sozialen Netzwerken funktionieren.

Von zwei ExpertInnen wurde zusätzlich erwähnt, dass es nicht die eine Lösung beziehungsweise ein perfektes Gesetz gibt, sondern dass es wichtig ist den Nutzen und den Schaden abzuwägen und dann zu entscheiden, was für eine Art von Regulierung am effektivsten wäre.

### 7.5.3 Forschungsfrage 3

#### FF 3: Wieso werden diese Angelegenheiten nicht auf EU-Ebene geregelt?

Der Verhaltenskodex, der auf EU-Ebene beschlossen wurde, ist zwar ein Ansatz, jedoch können die sozialen Netzwerke hier jederzeit wieder aussteigen. Aus diesem Grund stellt sich die Frage, wieso kein strengeres Gesetz auf EU-Ebene zur Regulierung von Hate Speeches beschlossen wird. Dies wurde unter anderem damit beantwortet, dass es in den verschiedenen EU-Staaten zu viele unterschiedliche Rechtssysteme gibt. Des Weiteren hat dies auch mit der historischen Entwicklung eines Landes zu tun. Die Länder Österreich und Deutschland fassen das Verhetzungsverbot beispielsweise sehr eng auf Grund der Geschichte des Holocausts. Aus diesem Grund sind diese beiden Länder auch daran interessiert nationale Gesetze einzuführen, die zur Bekämpfung von Hate Speeches beitragen sollen.

Ebenfalls wurde erwähnt, dass analysiert werden muss, wer Interesse an einer Regulierung auf EU-Ebene hätte und wer nicht und ob dies für EU-Staaten überhaupt umsetzbar wäre.

Obwohl eine EU-weite Lösung zur Bekämpfung gegen Hate Speeches bei zwei ExpertInnen auf Zuspruch stößt, wäre eine weltweite Lösung am sinnvollsten, denn das Internet hört nicht bei den EU-Grenzen auf. Zudem handelt es sich bei den sozialen Netzwerken wie Facebook, Twitter und Co. um global-agierende Konzerne, die auch am besten global reguliert werden sollten und nicht nur

innerhalb der EU oder auf nationaler Ebene. Dies würde jedoch eine internationale Zusammenarbeit erfordern.

#### 7.5.4 Forschungsfrage 4

##### **FF 4: Welche Rolle spielt die politische Situation in den einzelnen Ländern, wenn es um die Regulierung von sozialen Netzwerken geht?**

Bei dieser Frage waren sich die ExpertInnen ebenfalls darüber einig, dass die politische Situation Einfluss auf die Regulierung von sozialen Netzwerken zur Bekämpfung von Hate Speeches hat. Wenn beispielsweise ein bestimmtes Gesetz in Planung ist, dann hat hier die Politik sehr viel Einfluss darauf, wie dieses ausgestaltet und ausgerichtet werden soll. Einige ExpertInnen erwähnten auch, dass vor allem konservative und rechtsextreme Parteien gegen eine solche Regulierung sind. Haben solche Parteien also politischen Einfluss in einem Land, dann hat es auf Einfluss auf eine solche Regulierung.

Ebenfalls Einfluss auf diese Regulierung hat das Klima in einer Gesellschaft. Ist Rassismus beispielsweise sehr anerkannt oder verwenden Parteien in einem Land selbst rassistische Äußerungen, so kommt es hier zu einer Normalisierung. Dadurch wird der Bedarf sowie die politische Anstrengung, Hate Speeches zu regulieren, geringer.

Eine effektive Maßnahme zur Regulierung von Hate Speeches ist der Einsatz von Bildungsarbeit um die Medienkompetenz sowie die Zivilcourage zu steigern. Bestimmte Meldestellen könnten dafür sehr nützlich sein. Generell benötigt es Maßnahmen die rundum gesetzlicher Verordnungen notwendig wären. Dafür werden jedoch finanzielle Förderungen benötigt. Von einer Expertin wurde dies bezüglich erwähnt, dass in Österreich unter der bisherigen Regierung sehr wenig passiert ist. Das heißt, die nationale Politik hat auch sehr viel Einfluss darauf was außerhalb gesetzlicher Verordnungen gegen Hate Speeches getan werden kann. Erwähnt wurde hierbei jedoch auch, dass nicht nur die Politik Einfluss auf die Regulierung hat, sondern beispielsweise auch die „quasi“-Monopolstellung von Facebook. Es ist auch wichtig darauf zu achten, denn die Datafication spielt bei der Regulierung von sozialen Netzwerken ebenfalls eine zentrale Rolle.

## 7.6 Hypothesengenerierung

Auf Grund der Durchführung der empirischen Erhebung mittels ExpertInneninterviews sowie der qualitativen Inhaltsanalyse nach Mayring zur Auswertung können nun folgende Hypothesen aufgestellt werden. Diese können für weitere Forschungsarbeiten herangezogen werden.

H1: Die Erkennung von Hate Speech ist am effektivsten, wenn eine menschliche Moderation eingesetzt wird.

H2: Werden mehr Maßnahmen außerhalb gesetzlicher Verordnungen gegen Hate Speech getroffen, so verringert sich die Anzahl von Hassreden in den sozialen Netzwerken.

H3: Je internationaler Hate Speeches reguliert werden, umso effektiver ist die Bekämpfung von Hassreden im Internet.

H4: Je konservativer beziehungsweise rechtsextremer eine Regierung einem Land ist, desto geringer ist die Wahrscheinlichkeit, dass es zu einer Regulierung von Hate Speeches kommt.

## 8 Conclusio

Das letzte Kapitel dieser Forschungsarbeit befasst sich zu Beginn mit der Limitation. Hierbei wird beschrieben, was bei dieser Arbeit als kritisch betrachtet werden muss. Abschließend folgt darauf ein Fazit, das die Leitfrage dieser Arbeit beantwortet sowie ein Forschungsausblick.

### 8.1 Limitation

Zur Beantwortung der Leitfrage sowie der untergeordneten Forschungsfragen dieser Forschungsarbeit hat sich die Methode der ExpertInneninterviews als am geeigneten erwiesen. Eine der Schwierigkeiten bei dieser Methode war die Erreichbarkeit beziehungsweise die Kontaktierung der ExpertInnen. Während der Recherche nach passenden ExpertInnen wurde es häufig zum Problem, dass keine Kontaktdaten gefunden wurden. Zumeist mussten jene ExpertInnen deshalb über die Adresse der jeweiligen Organisation kontaktiert werden. Dies führte dazu, dass lange auf eine Rückmeldung gewartet werden musste. Die meisten E-Mails, die an eine allgemeine Firmenadresse gesendet wurden, wurden nicht beantwortet. Daher mussten diese ExpertInnen telefonisch kontaktiert werden. Insgesamt nahm die Kontaktierung sehr viel Zeit in Anspruch, so dass sich der gewünschte Zeitplan verschoben hat.

Obwohl die Grundgesamtheit auf Grund der definierten Kriterien relativ groß war, war es im gegebenen Forschungsrahmen lediglich möglich sechs ExpertInnen zu interviewen. Dies war deshalb der Fall, da sehr viele ExpertInnen absagen mussten beziehungsweise sich nicht zurückmeldeten. Da das sechste Interview schon relativ spät stattgefunden hat, wurden daraufhin aus zeitlichen Gründen keine weiteren ExpertInnen mehr befragt. Alle geführten Interviews können als sehr aussagekräftig beschrieben werden. Dennoch wäre die Meinung von JuristInnen zur Beantwortung dieser Leitfrage ebenfalls interessant gewesen und hätte mehr Einblicke in die Möglichkeiten der Regulierung von sozialen Netzwerken gegeben.

Wie schon erwähnt wäre es ideal ExpertInneninterviews face-to-face durchzuführen, da so eine bessere Interviewatmosphäre aufgebaut werden kann und die Gefahr von Informationsverlust geringer ist. Obwohl das face-to-face Interview hauptsächlich Vorteile aufweist war es der Autorin dieser Forschungsarbeit wegen gesundheitlicher Gründe zu dem Zeitpunkt der Interviewtermine nicht möglich eine längere Reise auf sich zu nehmen. Daher mussten alle sechs Interviews per Telefon durchgeführt werden.

## 8.2 Fazit

Ziel dieser Arbeit war es herauszufinden, wie soziale Netzwerke momentan reguliert werden beziehungsweise selber agieren um Hate Speeches zu bekämpfen. Außerdem sollten Informationen darüber gesammelt werden, welche Maßnahmen getroffen werden können um effektiv gegen Hassreden vorzugehen.

In der theoretischen Forschung dieser Arbeit konnten bereits vorhandene Informationen zu diesem Thema gesammelt werden. Bezuglich dieser Literaturrecherche kann zusammenfassend gesagt werden, dass soziale Netzwerke heute einen großen Einfluss auf die Kommunikation sowie die Meinungsbildung in unserer Gesellschaft haben. Vor allem Hate Speeches stellen dabei ein großes Problem dar. Um Hassreden in den sozialen Netzwerken zu vermeiden wurde innerhalb der EU ein Verhaltenskodex erstellt, den die sozialen Netzwerke unterschreiben konnten, jedoch nicht mussten. Dabei werden die sozialen Netzwerke dazu verpflichtet rechtswidrige Inhalte zu löschen. Deutschland hat daraufhin ebenfalls ein Gesetz zur Regulierung von Hate Speeches beschlossen und in Österreich ist ebenfalls ein Gesetz in Planung. Obwohl dadurch mehr Aufmerksamkeit auf dieses Thema fiel, wird viel Kritik an diesen Gesetzen ausgeübt. Daher hat sich folgende Leitfrage dieser Arbeit ergeben: „*Wie können soziale Netzwerke zukünftig reguliert werden, um Hate Speeches zu vermeiden?*“ Aus dieser Leitfrage haben sich vier untergeordnete Forschungsfragen ergeben, die daraufhin empirisch untersucht wurden.

Grundsätzlich kann gesagt werden, dass Hate Speech als ein sehr starkes Problem wahrgenommen wird. Es schädigt den öffentlichen Diskurs und führt zu einer Einschränkung der Meinungsfreiheit. Ein essenzielles Problem hinsichtlich der Regulierung ist hingegen, dass keine eindeutige Definition von Hate Speech existiert. Dies bedeutet, dass momentan versucht wird etwas zu regulieren, ohne zu wissen was reguliert werden soll. Dies stellt ein zentrales Problem für die Regulierung dar.

Trotz dieser nicht klaren Definition gibt es schon Gesetze wie das Netzwerkdurchsetzungsgesetz in Deutschland. Obwohl solche Regulierungen hilfreich sein können hat die Forschung ergeben, dass nicht nur auf einer Ebene reguliert werden sollte. Es ist notwendig nicht nur auf der Ebene der NutzerInnen, sondern auch auf der Ebene der sozialen Netzwerke zu regulieren. Das heißt, dass unter anderem auch auf die Struktur von den sozialen Netzwerken eingegangen werden soll. Es muss verstanden werden, wie sozialen Netzwerke funktionieren, so dass die Regulierungen dann auch Sinn ergeben. Des Weiteren ist man der Ansicht, dass global-agierende Unternehmen nicht auf nationaler Ebene reguliert werden sollten, sondern eine weltweite Lösung am sinnvollsten wäre.

Die Forschung hat außerdem ergeben, dass Artificial Intelligence als Regulierungsmethode noch zu wenig fortgeschritten ist. Auf Grund der Tatsache, dass AI den Kontext von gewissen Postings noch nicht versteht, ist es notwendig innerhalb der Plattformen mehr ExpertInnen einzustellen, die Hate Speech auch als solche erkennen. Momentan wird die Ansicht geteilt, dass es noch mehr menschliche Moderation benötigt um Hate Speeches zu identifizieren und zu entfernen. Als Idee wurde hier auch genannt ein unabhängiges Gremium aufzubauen, dass noch einmal prüfen soll, ob es sich bei gewissen Postings auch tatsächlich um eine Hassrede handelt.

Grundsätzlich bedarf es jedoch auch viele Maßnahmen rund um die gesetzlichen Verordnungen. Es ist wichtig am Kern des Problems anzusetzen und die Frage aufzustellen, warum Hass verbreitet wird. Daher kam die Forschung zu dem

Ergebnis, dass es viel Bildungsarbeit benötigt. Es wird angenommen, dass es bei vielen UserInnen an Medienkompetenz und Zivilcourage fehlt. Aus diesem Grund wird hier empfohlen, dass Schulen eine gewisse Medienbildung leisten sollten. Zudem sollen auch Melde- beziehungsweise Beratungsstellen mehr gefördert werden, um Bildungs- und Bewusstseinsarbeit bei Kindern und Jugendlichen leisten zu können.

Zusammenfassend kann gesagt werden, dass es bei der Regulierung vorrangig darum geht zu wissen, was reguliert werden soll. Außerdem muss an mehr Stellen angesetzt werden. Eine einfache gesetzliche Verordnung reicht nicht. Es ist wichtig sich dabei die Frage zu stellen, woher Hate Speeches kommen und wie hier am besten angesetzt werden kann. Zudem sollten auch Plattformunternehmen mehr in die Erkennung von Hate Speech investieren. Bei der Regulierung von sozialen Netzwerken zur Vermeidung von Hate Speech gibt es grundsätzlich kein perfektes Gesetz. Es müssen immer die Vor- und die Nachteile abgewogen werden. Hauptsächlich bedarf es bei so einer Regulierung jedoch eine enge Zusammenarbeit aller Akteure: Politik, Beratungsstellen und Plattformunternehmen.

### **8.3 Ausblick**

Es wurde mit dieser Forschungsarbeit ermittelt, wie am effektivsten gegen Hassreden im Internet vorgegangen werden kann. Dieses Thema bietet jedoch die Möglichkeit noch weitere Forschungen anzustellen. Es kann beispielsweise noch mehr in die Tiefe gegangen werden, vor allem wenn es um juristische Fragestellungen geht. Des Weiteren wäre es interessant diese Forschung geografisch auszuweiten und auch mit ExpertInnen außerhalb Europas zu sprechen. Interessant wäre es ebenfalls PolitikerInnen in so ein Forschungsprojekt miteinzubeziehen. Außerdem wäre es spannend zu ermitteln, inwiefern sich die Ergebnisse dieser Arbeit durch eine andere empirische Untersuchung bestätigen lassen. Die fünf abgeleiteten Hypothesen dieser Forschungsarbeit könnten dabei richtungsweisend sein.

# Literaturverzeichnis

- Berka, W., & Trappel, J. (2019). *Internetfreiheit in Österreich*. Wien: Manz.
- BMSGK. (2019). Gewalt im Netz. Abgerufen 11. Juli 2019, von <http://www.bmwf.gv.at/cms/home/thema.html?channel=CH1631&doc=CMS1484137199089>
- Bogner, A., Littig, B., & Menz, W. (2014). *Interviews mit Experten. Eine praxisorientierte Einführung*. Wiesbaden: Springer Fachmedien.
- Bogner, A., & Menz, W. (2009). Experteninterview in der qualitativen Sozialforschung. Zur Einführung in eine sich intensivierende Methodendebatte. In A. Bogner, B. Littig, & W. Menz, *Experteninterview. Theorien, Methoden, Anwendungsfelder* (S. 7–34). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brodnig, I. (2016). *Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können*. Christian Brandstätter Verlag.
- Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), 297–326. <https://doi.org/10.1177/1468796817709846>
- Conway, R. S., Danielson, J. T., & Stryker, R. (2016). What is political incivility. *Communication Monographs*, 83(4), 535–556. <https://doi.org/10.1080/03637751.2016.1201207>
- Delgado, R. (2018). Words that wound: A tort action for racial insults, epithets and name calling. In R. Delgado, C. R. Lawrence, M. J. Matsuda, & K. Williams Crenshaw (Hrsg.), *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (S. 89–110). New York: Routledge.

Demokratiezentrum Wien. (2019). Maßnahmen auf politischer Ebene. Abgerufen 18. Juli 2019, von <http://www.demokratiezentrum.org/themen/mediengesellschaft/virtuelle-agora-und-digitale-zivilcourage/massnahmen-auf-politischer-ebene.html>

Der Standard. (2018). Neue Antworten auf Hate-Speech. Abgerufen 2. Juli 2019, von <https://www.derstandard.at/story/2000090293041/neue-antworten-auf-hate-speech>

Der Standard. (2019a). Hassposting gegen Neujahrssbaby: Pensionist verurteilt. Abgerufen 15. Juli 2019, von <https://www.derstandard.at/story/2000099347070/hassposting-gegen-neujahrssbaby-pensionist-verurteilt>

Der Standard. (2019b). Regierung will Anonymität im Netz abschaffen – mit hohen Strafen. Abgerufen 27. Mai 2019, von <https://derstandard.at/2000101143621/Regierung-will-Anonymitaet-im-Netz-abschaffen-mit-hohen-Strafen>

Deussen, O., Krafft, T. D., & Zweig, K. A. (2017). Algorithmen und Meinungsbildung. *Informatik Spektrum*, 40(4), 318–326. <https://doi.org/10.1007/s00287-017-1053-2>

Deutscher Bundestag. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz -NetzDG)*. (2017).

Deutscher Bundestag. (2018). *Regulierung von Hate Speech und Fake News in sozialen Netzwerken durch EU-Mitgliedsstaaten*.

Die Presse. (2015). FPÖ-Politiker wegen Hasspostings aus Partei ausgeschlossen. Abgerufen 3. Juli 2019, von <https://diepresse.com/home/politik/innenpolitik/4801228/FPOePolitiker-wegen-Hasspostings-aus-Partei-ausgeschlossen>

Die Presse. (2019). Hassposting über Flüchtlinge: Niederösterreicher bekommt Diversion. Abgerufen 13. Juni 2019, von [https://diepresse.com/home/panorama/oesterreich/5558575/Hassposting-ueber-Fluechtlinge\\_Niederoesterreicher-bekommt-Diversion](https://diepresse.com/home/panorama/oesterreich/5558575/Hassposting-ueber-Fluechtlinge_Niederoesterreicher-bekommt-Diversion)

DIVSI. (2013). *Freiheit versus Regulierung im Internet*. Hamburg.

Eckes, C., Fernholz, T., Geschke, D., Klaßen, A., & Quent, M. (2018). *Hass im Netz: Der schleichende Angriff auf unsere Demokratie* (Institut für Demokratie und Zivilgesellschaft, Hrsg.). Jena.

Europäisches Parlament. *Richtlinie 2000/31/EG des Europäischen Parlaments und des Rates*. (2000).

European Commission. (2019). The EU Code of conduct on countering illegal hate speech online. Abgerufen 18. Juli 2019, von [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en)

Fanta, A. (2019). Digitales Vermummungsverbot: Österreich will Klarnamen und Wohnsitz von Forennutzern. Abgerufen 23. August 2019, von <https://netzpolitik.org/2019/digitales-vermummungsverbot-oesterreich-will-klarnamen-und-wohnsitz-von-forennutzern/>

Flick, U. (2007). *Qualitative Sozialforschung. Eine Einführung*. Rheinbek bei Hamburg: Rowohlt Taschenbuch Verlag.

Forsa. (2018). *Hate-Speech Ergebnisbericht*. Düsseldorf.

Fox, B. (2018). Gesetze gegen „Hassrede“ vs. Meinungsfreiheit. Abgerufen 4. August 2019, von <https://www.euractiv.de/section/digitale-agenda/news/gesetze-gegen-hassrede-vs-meinungsfreiheit/>

Fumagalli, C. (2019). Propositional attitudes, harm and public hate speech situations: Towards a maieutic approach. *European Journal of Political Theory*, 0(0), 1–22. <https://doi.org/10.1177/1474885119836627>

Ganz, K. (2019). Hate Speech im Internet. In J. Dorer, B. Geiger, B. Hipfl, & V. Ratkovic (Hrsg.), *Handbuch Medien und Geschlecht* (S. 1–10). Wiesbaden: Springer VS.

Gläser, J., & Laudel, G. (2009). *Experteninterview und qualitative Inhaltsanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Haenlein, M., & Kaplan, A. M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53, 59–68.

Handelsblatt. (2019). Dauerkritik: Bekämpfung von Hassiraden – Facebook plant Aufsichtsgremium. Abgerufen 19. Juli 2019, von <https://www.handelsblatt.com/technik/it-internet/dauerkritik-bekaempfung-von-hassiraden-facebook-plant-aufsichtsgremium-/24489506.html>

Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242. <https://doi.org/10.1515/jplr.2010.011>

Heise, N. (2016). Algorithmen. In J. Heesen (Hrsg.), *Handbuch Medien- und Informationsethik*. (S. 202–209). Stuttgart: Metzler.

Hoppenstedt, M. (2017). Zu Besuch in Facebooks neuem Löschzentrum, das gerade den Betrieb aufnimmt. Abgerufen 19. Juli 2019, von <https://www.vice.com/de/article/qv37dv/zu-besuch-in-facebook-s-neuem-loeschzentrum-das-gerade-den-betrieb-aufnimmt>

Humborg, C., & Nguyen, T. A. (2018). *Die publizistische Gesellschaft. Journalismus und Medien im Zeitalter des Plattformkapitalismus.* Wiesbaden: Springer VS.

ISA. (2017). *Digitalmonitor.* Wien.

Jourová, V. (2016). *Code of conduct on countering illegal hate speech online: First results on implementation.*

Jourová, V. (2019). *Code of Conduct on countering illegal hate speech online.*

Kleine Zeitung. (2018). Twitter: Mit künstlicher Intelligenz gegen Trolle im Internet. Abgerufen 19. Juli 2019, von <https://www.kleinezeitung.at/wirtschaft/5454560/>

Kolo, C. (2016). Social Media. In J. Krone & T. Pellegrini (Hrsg.), *Handbuch Medienökonomie. Springer Nachschlagewissen.* (S. 2–18). Wiesbaden: Springer VS.

Krüger, U. (2018). Der neue Strukturwandel der Öffentlichkeit und die German Angst. In G. Hooffacker, W. Kenntemich, & U. Kulisch (Hrsg.), *Die neue Öffentlichkeit. Wie Bots, Bürger und Big Data den Journalismus verändern* (S. 9–26). Wiesbaden: VS Verlag.

Kurier. (2018). Hass im Netz richtet sich vor allem gegen Moslems und Flüchtlinge. Abgerufen 13. Juni 2019, von <https://kurier.at/chronik/oesterreich/hass-im-netz-richtet-sich-vor-allem-gegen-moslems-und-fluechtlinge/312.627.923>

Kurz, C. (2018). Die Meinungsfreiheit und das NetzDG: Schwerwiegender Verstoß gegen Grundrecht. Abgerufen 4. August 2019, von <https://netzpolitik.org/2018/die-meinungsfreiheit-und-das-netzdg-schwerwiegender-verstoss-gegen-grundrecht/>

Langton, R. (2012). Beyond belief: Pragmatics in hate speech and pornography. In I. Maitra & M. K. McGowan (Hrsg.), *Speech and Harm: Controversies over Free Speech* (S. 72–90). Oxford: Oxford University Press.

Latzer, M., Just, N., Saurwein, F., & Slominski, P. (2002). *Selbst- und Ko-Regulierung im Mediamatiksektor. Alternative Regulierungsformen zwischen Staat und Markt*. Wiesbaden: Springer Verlag.

Lischka, K., & Stöcker, C. (2017). *Digitale Öffentlichkeit. Wie algorithmische Prozesse den gesellschaftlichen Diskurs beeinflussen*. Gütersloh: Bertelsmann Stiftung.

Löblich, M., Meyen, M., Pfaff-Rüdiger, S., & Riesmeyer, C. (2011). *Qualitative Forschung in der Kommunikationswissenschaft. Eine praxisorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Mack, D., & Vilberger, D. (2016). *Social Media für KMU. Der Leitfaden mit allen Grundlagen Strategien und Instrumenten*. Wiesbaden: Springer Gabler.

Mahnke, M. (2015). Der Algorithmus, bei dem man mit muss? Ein Perspektivwechsel. *Communicatio Socialis*, 48(1), 34–45.

Matsuda, M. J. (2018). Public response to racist speech: Considerin the victim's story. In R. Delgado, C. R. Lawrence, M. J. Matsuda, & K. Williams Crenshaw (Hrsg.), *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (S. 17–50). New York: Routledge.

May, M. (2018). Hate Speech analog – Eine situative Herausforderung in Schule und Unterricht. *GWP - Gesellschaft. Wirtschaft. Politik*, 67(3), 399–408. <https://doi.org/10.3224/gwp.v67i3.12>

Mayring, P. (2010). Design. In G. Mey & K. Mruck (Hrsg.), *Qualitative Forschung in der Psychologie* (S. 225–238). Wiesbaden: Springer Fachmedien.

Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim/Basel: Beltz Verlag.

Meier, W. A. (2011). Demokratie und Media Governance in Europa. Regulierung, Partizipation, Mitbestimmung. In H. J. Kleinstuber & S. Nehls (Hrsg.), *Media Governance in Europa* (S. 37–57). Wiesbaden: VS Verlag.

Meuser, M., & Nagel, U. (2009). Experteninterview und der Wandel der Wissensproduktion. In A. Bogner, B. Littig, & W. Menz, *Experteninterview. Theorien, Methoden, Anwendungsfelder* (S. 35–60). Wiesbaden: VS Verlag für Sozialwissenschaften.

Mey, G., & Mruck, K. (2010). *Qualitative Forschung in der Psychologie*. Wiesbaden: Springer Fachmedien.

Mühlenmeier, L. (2017). Twitter plant, Hate Speech per Algorithmus zu erkennen. Abgerufen 22. August 2019, von Netzpolitik.org website: <https://netzpolitik.org/2017/twitter-will-hate-speech-per-algorithmus-erkennen/>

Niklewicz, K. (2017). Weeding Out Fake News: An Approach to Social Media Regulation: *European View*, 29–30. <https://doi.org/10.1007/s12290-017-0468-0>

Österreichisches Parlament. (2019). Bundesgesetz über Sorgfalt und Verantwortung im Netz; KommAustria-Gesetz, Änderung. Abgerufen 31. 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)

Puppis, M. (2010). Media Governance: A New Concept for the Analysis of Media Policy and Regulation. *Culture & Critique*, 3, 134–149. <https://doi.org/10.1111/j.1753-9137.2010.01063.x>

Rawls, J. (1999). *A Theory of Justice: Revised Edition*. Cambridge: Harvard University Press.

Rondinella, G. (2019). Kampf gegen Hass und Gewalt: Facebook will stärker in Künstliche Intelligenz investieren. Abgerufen 19. Juli 2019, von <https://www.horizont.net/tech/nachrichten/kampf-gegen-hass-und-gewalt-facebook-will-staerker-in-kuenstliche-intelligenz-investieren-174564>

Roßmann, R. (2019). Facebook muss Bußgeld in Millionenhöhe zahlen. Abgerufen 3. August 2019, von <https://www.sueddeutsche.de/digital/facebook-strafe-hasskommentare-netzdg-bundesamt-fuer-justiz-1.4508415>

Rozgonyi, K. (2018). A New Model for Media Regulation. *Intermedia*, 46(1), 18–23.

Saferinternet.at. (2016). Hasspostings im Internet – was sagt das Gesetz? Abgerufen 31. Juli 2019, von <https://www.saferinternet.at/news-detail/hasspostings-im-internet-was-sagt-das-gesetz/>

Saurwein, F. (2019). *Algorithmen als Herausforderung für Governance*. Gehalten auf der Symposium Medienethik, St. Pölten.

Schmid, F. (2016). „Es reicht“: Grüne verklagen Facebook wegen Hasspostings. Abgerufen 7. Juni 2019, von <https://derstandard.at/2000044481423/Es-reicht-Gruene-verklagen-Facebook-wegen-Hasspostings>

Schmid, F. (2018). Glawischnig-Klage: EuGH soll prüfen, ob Facebook Hasspostings selbst suchen muss. Abgerufen 7. Juni 2019, von <https://derstandard.at/2000071915249/Glawischnig-Klage-EuGH-soll-pruefen-ob-Facebook-Hasspostings-selbst-suchen>

Schmidt, J.-H. (2018). *Social Media* (2. Aufl.). Wiesbaden: Springer VS.

Schrape, J.-F., & Siri, J. (2019). Facebook und andere soziale Medien. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (2. Aufl., S. 1053–1064). Wiesbaden: Springer VS.

Schweiger, W. (2017). *Der (des)informierte Bürger im Netz. Wie soziale Medien die Meinungsbildung verändern*. Wiesbaden: Springer Fachmedien.

Seglow, J. (2016). Hate speech, dignity and self respect. *Ethical Theory and Moral Practice*, 19(5), 1103–1116. <https://doi.org/10.1007/s10677-016-9744-3>

Seifter, T. (2017). Hasskommentare im Internet. Die Identitäre Bewegung im Spiegel von Facebook-Kommentaren im ORF. *Grazer Linguistische Studien*, 88, 47–85. <https://doi.org/10.25364/04.44:2017.88.3>

Simpson, R. M. (2013). Dignity, Harm and Hate Speech. *Law and Philosophy*, 32(6), 701–728. <https://doi.org/10.1007/s10982-012-9164-z>

Sponholz, L. (2018). *Hate Speech in den Massenmedien. Theoretische Grundlagen und empirische Umsetzung*. Wiesbaden: Springer VS.

Stegler, J. (2018). Facebook, Twitter, Instagram: Wie soziale Netzwerke gegen den Hass kämpfen. Abgerufen 19. Juli 2019, von <https://www.handelsblatt.com/unternehmen/it-medien/facebook-twitter-instagram-wie-soziale-netzwerke-gegen-den-hass-kaempfen/22781312.html>

Süddeutsche Zeitung. (2018). Das Netz der schwarzen Listen. Abgerufen 19. Juli 2019, von <https://www.sueddeutsche.de/digital/upload-filter-das-netz-der-schwarzen-listen-1.3835794-2>

Tribe, B. (2019). Internet-Regulierung: Großbritannien will alle Firmen überwachen. Abgerufen 4. August 2019, von <https://www.nzz.ch/wirtschaft/internet-regulierung-grossbritannien-will-alle-firmen-ueberwachen-ld.1473610>

Turner, E. (2017). Hate Speech. Das Subjekt des Widerstands. In T. Spies & E. Tuider (Hrsg.), *Biographie und Diskurs. Methodisches Vorgehen und Methodologische Verbindungen* (S. 111–127). Wiesbaden: VS Verlag.

Waldron, J. (2012). *The Harm in Hate Speech*. Cambridge/London: Harvard University Press.

Wettstein, M. (2016). Ursache und Umgang mit Hass-Kommentaren im Internet. In Parlamentsdirektion Österreich (Hrsg.), *Grünbuch Digital Courage*.

Whitten, S. (2019). Recognition, Authority Relations, and Rejecting Hate Speech. *Ethical Theory and Moral Practice*, 1–17. <https://doi.org/10.1007/s10677-019-10003-z>

ZARA. (2018). *Rassismus Report 2018*. Wien.

# Anhang

## Exposé Master These

## 1. Abgabe

Familienname	<b>Berke</b>	
Vorname	Lisa	
Matrikelnummer	mm171803	
Mailadresse	mm171803@fhstp.ac.at	
Telefonnummer	0676/5073574	
Datum	24.01.2019	
Name Betreuer	FH-Prof. Priv.-Doz. Dr. Michael Litschka	Bestätigung der Betreuung durch Unterschrift des Betreuers
Thema	Medienethik & Content Management	
Arbeitstitel	Hate-Speech in Zeiten von Social Media - Wie können Soziale Netzwerke zukünftig reguliert werden um Hassreden und Hast-Postings zu vermeiden?	
Problemstellung/Forschungsfrage	Es war noch nie möglich so einfach, ohne Aufwand schnell Informationen zu teilen und miteinander zu kommunizieren und diskutieren. Seit des Aufkommens der modernen Online-Kommunikation, die durch das Web 2.0 ermöglicht wurde, spricht man von einer veränderten medialen Öffentlichkeit.	

Facebook, Twitter oder YouTube als soziale Netzwerke sind heute in der täglichen Kommunikation eine Notwendigkeit. Besonders an diesen Plattformen ist, dass jeder die Möglichkeit hat aktiv und selbstbestimmt in das Kommunikationsgeschehen einzugreifen. Dabei kann persönlich zum Sender avanciert werden und sich ungefiltert und uneingeschränkt der ganzen Welt mittzueilen. (vgl. Seifter 2017, S. 54)

Die Online-Kommunikation hat jedoch nicht nur positive Seiten, sondern weist auch negative Seiten auf, wie etwa beleidigende Kommentare auf Webseiten von Zeitungen auf Facebook oder Twitter. (vgl. ebd., S. 55)

Faktoren, die Hasspostings innerhalb der Online-Kommunikation bewirken, sind hauptsächlich die Anonymität und die physische Distanz bzw. Unsichtbarkeit. Diese Anonymität im Internet aufzuheben ist jedoch widersprüchlich. Zum einen wäre eine Registrierungspflicht- bzw. ein Authentifizierungszwang mit einer Art Einschränkung der Meinungsfreiheit gleichzusetzen, da die Bereitschaft zur Partizipation wegen sozialer Ängste (Verlust des Arbeitsplatzes, gesellschaftliche Isolation..) bei kritischer Meinungsäußerung nachvollziehbar ist. Zum anderen konnte bis jetzt nicht erwiesen werden, dass das Internet durch seine Konstitution, vor allem durch die Anonymität, Hassreden verursacht. (vgl. ebd., S. 57)

Nicht zu leugnen ist jedoch, dass Hassreden auch im Internet existieren. Beispielsweise sind diese auf Facebook zum Beispiel auf diversen Facebook-

Seiten und -Gruppen zu finden. Plattformen wie Facebook sind wegen der Möglichkeit der Personalisierung des Kontos auch bekannt dafür, eine Verstärkung der eigenen Meinung zu bewirken. Dies passiert deshalb, da Facebook-Angebote gezielt von ähnlich gesinnten Personen aufgesucht werden, das dazu führt, dass kaum inhaltliche Abwechslung oder argumentative Ausgewogenheit des Kommunikationsraums vorhanden ist. Bei diesem Phänomen spricht man von sogenannten „Echokammern“. Durch diese Echokammern kommen „Teilöffentlichkeiten“ zu Stande. Dadurch kommt der Gesellschaft die gemeinsame Basis für einen politischen Diskurs abhanden und ,so dass eine kollektive Meinungs- und Willensbildung erschwert wird. Dabei kommt es zu Hass im Netz, der politische und gesetzliche Relevanz hervorruft. (vgl. ebd., S. 58)

Die zukünftige Regulation von Online-Medien bezüglich Hassreden, der Schutz von Minderheiten, die virale Verbreitung von Fake-News auf den Sozialen Medien und den Kampf um das Urheberrecht haben hitzige Debatten in der Europäischen Kommission ausgelöst. Innerhalb der Politiker in Europa ist man sich einig, dass man schnell zu Lösungen kommen muss um diese Phänomene zu stoppen. (vgl. Rozgonyi 2018, S. 18)

Im Mai 2016 haben die Unternehmen Facebook, Twitter, Youtube und Microsoft freiwillig einem Verhaltenskodex zugestimmt. Der Zweck dieses Kodex war es besser gegen Hassreden vorzugehen und die „notice and take down“ Regel effektiver zu

machen. Die Plattformen haben sich dabei verpflichtet, illegale Inhalte innerhalb von 24 Stunden zu löschen. Wird dies nicht eingehalten so können diese Plattformen mit einer Geldstrafe mit bis zu 50 Millionen Euro belangt werden. Die erwartende Wirkung erreichte dieser Kodex jedoch nicht. Obwohl solche illegalen Inhalte meistens sofort entfernt wurden, wurden diese vor der Entfernung schon so oft weiterverbreitet so dass es viele andere Nutzer ebenfalls posten können. Diese weiterverbreiteten Inhalte werden jedoch ignoriert. Es kann also gesagt werden, dass die bisherigen Bemühungen Soziale Netzwerke zu regulieren gescheitert sind. (Vgl. Niklewicz 2017, S. 29-30)

Aus diesem Grund stellt sich die Frage wie man heute am effektivsten gegen Hassreden und Hass-Postings auf Sozialen-Medien vorgehen kann.

Leitfrage: Wie können Soziale Netzwerke zukünftig reguliert werden um Hassreden und Hass-Postings zu vermeiden?

FF 1: Welche Rolle könnten Techniken wie KI bei der Bekämpfung von Hassreden und Hass-Posting in Sozialen Netzwerken spielen?

FF2: Wie effizient wäre eine Identifikationspflicht um gegen Hass im Netz vorzugehen?

FF3: Inwiefern würde eine Auskunftspflicht über Hasspostings und Hassreden von Usern von Sozialen Netzwerken an die Justiz Sinn machen?

	<p>FF4: Wieso regelt man diese Angelegenheit nicht auf internationaler Ebene?</p> <p>FF5: Welche Rolle spielt die politische Situation in den einzelnen Ländern, wenn es um die Regulierung von Sozialen Netzwerken geht?</p>
Aufbau und Gliederung	<ol style="list-style-type: none"> <li>1. Einleitung             <ol style="list-style-type: none"> <li>1.1 Einführung in die Thematik</li> <li>1.2 Problematik und Zielsetzung</li> <li>1.3 Aufbau und Methodik der Arbeit</li> <li>1.4 Forschungsstand</li> </ol> </li> <li>2. Soziale Netzwerke             <ol style="list-style-type: none"> <li>2.1 Erläuterung</li> <li>2.2 Die Veränderung der Kommunikation</li> <li>2.3 Die Problematik der Echokammern</li> <li>2.4 Hass in den Sozialen Netzwerken</li> </ol> </li> <li>3. Das Soziale Netzwerk Facebook             <ol style="list-style-type: none"> <li>3.1 Erläuterung</li> <li>3.2 Die Nutzung von Facebook (in Ö)</li> <li>3.3 Hassreden &amp; Hass-Postings und deren Auswirkungen (Beispiele)</li> </ol> </li> <li>4. Medienregulierung in der EU             <ol style="list-style-type: none"> <li>4.1 Allgemein</li> <li>4.2 Regulierung unterschiedlicher Mediendienste                     <ol style="list-style-type: none"> <li>4.2.1 AVMD-RL</li> <li>4.2.2. E-Commerce Richtlinie</li> </ol> </li> <li>4.3. Der Versuch der Internetregulierung                     <ol style="list-style-type: none"> <li>4.3.1 Telekommunikationsrichtlinie</li> </ol> </li> </ol> </li> <li>5. Die Regulierung von Sozialen Netzwerken             <ol style="list-style-type: none"> <li>5.1 Allgemein</li> <li>5.2 Verschiedene Ansätze</li> <li>5.3 Die Problematik mit der Einschränkung der Meinungsfreiheit</li> </ol> </li> <li>6. Zwischenfazit &amp; Forschungsfragen</li> </ol>

	<p>7. Methodik</p> <p>7.1 Die Erläuterung der Erhebungsmethode</p> <p>7.2 Die Vorgangsweise</p> <p>7.3 Erläuterung der Auswertungsmethode</p> <p>8. Empirie</p> <p>8.1 Ergebnisanalyse</p> <p>8.2 Interpretation der Ergebnisse</p> <p>8.3 Beantwortung der Forschungsfragen</p> <p>8.4 Hypothesenbildung</p> <p>9. Conclusio</p> <p>9.1 Limitation</p> <p>9.2 Fazit</p> <p>Literaturverzeichnis</p> <p>Anhang</p>
Methodenwahl	<p>Um die Forschungsfragen dieser Arbeit beantworten zu können, wird hier das Fachwissen von ExpertInnen benötigt. Aus diesem Grund ist hier eine qualitative Methode, nämlich das Experteninterview, am besten geeignet.</p> <p>Die Befragungen der ExpertInnen werden als Leitfadeninterviews durchgeführt. Dabei wird mit offenen Fragen gearbeitet, so dass zwischen verschiedene Phasen und Bereiche abgegrenzt werden kann. (vgl. Harke, L'hoest, Wingen 2000, S. 14).</p> <p>Ausgewertet werden diese Interviews mittels der qualitativen Inhaltsanalyse nach Mayring.</p> <p>Um zu einer Beantwortung meiner Forschungsfragen zu kommen werden sieben bis neun ExpertInnen interviewt, die sich intensiv mit diesem Thema beschäftigen. Das könnten unter</p>

	<p>anderen JuristInnen sein, die sich laufend mit Cyber-Kriminalität beschäftigen oder sich generell mit dem Thema Recht im Internet befassen. Des Weiteren könnten auch JournalistInnen oder Autoren, die diese Thematik ständig beobachten, interviewt werden. Auch ExpertInnen aus der Politik kämen für diese Interviews in Frage.</p> <p>All diese ExpertInnen könnten Aufschlüsse darüber geben, wie sich Soziale Netzwerke in Zukunft am besten regulieren lassen würden.</p>
Literaturhinweise	<p>Banghart, Scott/Etter, Michael/Stohl, Cynthia (2018): Organizational Boundary Regulation Through Social Media Policies. In: Sage Journals, Vol. 32 (3), S. 337-373</p> <p>Banks, James (2010): Regulating Hate Speech Online. In: International Review of Law, Computers and Technology, Vol. 24 (3), S. 233-239</p> <p>Brodnig, Ingrid (2016): Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können. Brandstätter Verlag</p> <p>Foxman, Abraham H/Wolf, Christopher (2013): Viral Hate. New York: St. Martin´s Press</p> <p>Gläser, Jochen (2010): Experteninterviews und qualitative Inhaltsanalyse. Wiesbaden: VS Verlag</p> <p>Harke, Andreas; L'hoest Andrea; Wingen, Sascha (2000): Vorgehensweise bei einer qualitativen Expertenbefragung am Beispiel Fondsmanager.</p>

	<p>Bochum: Ruhr-Universität</p> <p>Hovy, Dirk/Waseem, Zeerak (2016): Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. Bei: NACCL-HLT, San Diego</p> <p>Mayring, Philipp (2015): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim und Basel: Beltz Verlag</p> <p>Niklewicz, Konrad (2017): Wedding Out Fake News. An Approach to Social Media Regulation. Brüssel: Wilfried Martens Centre for European Studies</p> <p>Rozgonyi, Krisztina (2018): A New Model For Media Regulation. In: InterMEDIA, Vol. 46 (1), S.18</p> <p>Seifert, Thorsten (2017): Hasskommentare im Internet – die identitäre Bewegung im Spiegel von Facebook-Kommentaren im ORF. Graz: Karl-Franzens-Universität</p> <p>Sheering, Cathal (2017): Regulations are a bigger threat than fake news. In: Media Development, Vol. 2, S. 33-35</p> <p>Sivek, Susan Currie (2010): Social Media Under Social Control: Regulating Social Media and the Future of Socialization. In: Sage Journals, Vol 4 (3), S. 146-164</p>
--	---

## Leitfaden der ExpertInneninterviews

Nur kurz als Einstiegsfrage: Wie empfinden Sie das Problem mit den Online Hate Speeches?

1. Die Soziale Medien haben heute eine große Marktmacht. Der Politik fällt es oft schwer, gegen Soziale Medien vorzugehen.  
Hinsichtlich Hate Speeches sagen Facebook, Twitter & YouTube, dass sie gegen Hass im Netz sind und versuchen alle Hass-Beiträge zu löschen → benutzen auch AI (Artificial Intelligence)
  - Wie sinnvoll ist AI als Methode?
  - Macht Artificial Intelligence da Sinn? → erkennen diese so etwas überhaupt? (wollen dies bezüglich nun auch mehr Personal einstellen)
  - Sollten Personen Hass-Postings nicht vorher überprüfen, bevor diese gelöscht werden? Denn AI stuft oft auch schon Satire als Hate Speech ein.
2. Ab 2020 gibt es in Österreich eine Auskunftspflicht (Anonymität wird aufgehoben; Plattformbetreiber müssen Vor- und Nachname inkl. Adresse des Nutzers im Hintergrund speichern) für Forennutzer von Plattformunternehmen. Was sagen Sie dazu?
  - Empfinden Sie diese Regelung als hilfreich? → geht darum, Täter schneller zur Rechenschaft zu ziehen
  - Glauben Sie, dass Hate Speeches dadurch weniger werden?
3. Was halten Sie von dem Netzwerkdurchsetzungsgesetz in DE (Betreiber sozialer Netzwerke wie Facebook, Twitter und YouTube müssen „offensichtlich rechtswidrige Inhalte innerhalb von 24 Stunden“ nach Eingang einer Beschwerde löschen oder sperren)
  - Finden Sie, dass dies eine bessere Lösung, als die die für Österreich vorgeschlagen wurde, wäre? Obwohl man auch hier Kritik übt, da dies ebenfalls die Meinungsfreiheit einschränkt?
4. Bezuglich der möglichen Regulierung. Momentan regelt dies ja jedes Land für sich. Wieso macht das nicht die EU?
  - Würde dies nicht viel sinnvoller sein?
  - Sind zu viele Länder dagegen wegen Einschränkung der Meinungsfreiheit?
  - Oder sind Sie der Meinung, dass eine Selbstregulierung effektiver sei? (Also ohne Einfluss des Staates, sondern durch eine Organisation, wie beispielsweise den Presserat) oder wäre auch Ko-Regulierung eine Lösung? (Bist jetzt ist es ja eher eine Selbstorganisation)
5. Glauben Sie, dass die politische Situation in einem Land Einfluss auf die Regulierung von Hate Speeches hat?
  - Sind bestimmte Parteien dagegen?
  - In den USA beispielsweise ist anhand des 1. Zusatzartikels der Verfassung verboten, Gesetze einzuführen, die gegen die Redefreiheit sind.
  - Finden Sie, dass dies die Redefreiheit beeinflusst?
  - Wo ist die Grenze zwischen Hate Speech und Free Speech?