Schriften zum Medien- und Informationsrecht 45

Steinebach | Bader | Rinsdorf | Krämer | Roßnagel (Hrsg.)

Desinformation aufdecken und bekämpfen

Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität

Nomos

Schriften zum Medien- und Informationsrecht

herausgegeben von
Prof. Dr. Boris P. Paal, M.Jur.

Band 45

Martin Steinebach | Katarina Bader | Lars Rinsdorf
Nicole Krämer | Alexander Roßnagel (Hrsg.)

Desinformation aufdecken und bekämpfen

Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität

Nomos

GEFÖRDERT VOM

Die Open Access-Veröffentlichung der elektronischen Ausgabe dieses Werkes wurde ermöglicht mit Unterstützung durch das Fraunhofer-Institut für Sichere Informationstechnologie (SIT) zur Förderung der wissenschaftlichen Forschung.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

1. Auflage 2020

Publiziert von
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3-5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3-5 | 76530 Baden-Baden

ISBN (Print): 978-3-8487-6390-0
ISBN (ePDF): 978-3-7489-0481-6

DOI: https://doi.org/10.5771/9783748904816

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1: Einleitung

Autorin:

Prof. Dr. Katarina Bader

Wurde durch die Digitalisierung ein postfaktisches Zeitalter eingeläutet? Nachdem im US-Präsidentschaftswahlkampf 2016 im Internet zahlreiche Falschmeldungen zirkulierten und dann ein Kandidat gewählt wurde, der immer wieder mit falschen Zahlen und unwahren Behauptungen Wahlkampf gemacht hatte, war die Aufregung auch außerhalb der USA groß. In Deutschland fürchteten viele, dass auch hierzulande Wahlen von Fake News[1], Malicious (also bösartigen) Bots und anderen Elementen digitaler Desinformationskampagnen bestimmt werden könnten. Im Vorfeld der Bundestagswahl 2017 und der Europawahl 2019 durchgeführte empirische Studien geben jedoch vorsichtig Entwarnung: Es wurden zwar sich viral verbreitende Falschmeldungen (Sängerlaub et al., 2018), sehr aktive Junk-News-Seiten (Marchal et al., 2019; Neudert et al., 2017), offensichtlich zusammengekaufte Follower[2] und Anzeichen für den Einsatz von Malicious Bots (Neudert et al., 2017: 4) gefunden, aber die Untersuchungen zeigen zugleich: Die Reichweite und Breitenwirkung digitaler Desinformationskampagnen ist in Deutschland noch vergleichsweise gering. Dabei ist auffällig, dass die meisten Studien den Hotspot der deutschsprachigen digitalen Desinformation im rechtspopulistischen Milieu ausfindig machten (Marchal et al., 2019: 6; Neudert et al., 2017: 4; Sängerlaub et al., 2018: 3). Also Entwarnung?

Sich beruhigt zurückzulehnen, wäre sicher falsch: Wahlkämpfe sind zwar neuralgische Punkte im demokratischen Prozess, dennoch wird man die Risiken, die digitale Desinformation in sich birgt, nicht vollständig verstehen, wenn man ausschließlich Wahlkampfphasen untersucht und sich dabei zudem auf das Erfassen der relativen Reichweite einer bestimmten, nachweislich erfundenen Meldung oder einzelner Junk-News-Seiten konzentriert. Vielmehr ist es notwendig, das Phänomen der digitalen Desinformation auch abseits

1 Begriffsbestimmungen in Kap. 1.C.

2 https://www.zdf.de/nachrichten/heute/us-studie-davis-verdaechtige-accounts-unterstuetzen-afd-soziale-netzwerke-zdfcheck-100.html (Stand: 24.6.2019).

von Wahlkämpfen interdisziplinär zu untersuchen, in enger Zusammenarbeit von Informatikern und Technikwissenschaftlern, Juristen, Medienpsychologen und Kommunikationswissenschaftlern. Vor welche Herausforderungen digitale Desinformation demokratische Gesellschaften auch jenseits von Wahlen stellt und warum diese Herausforderungen nur interdisziplinär bearbeitet werden können, lässt sich an einem exemplarischen Fall erklären, der im Sommer 2018 deutschlandweit hitzige Debatten auslöste.

A. *Die Wirkung von Desinformation – ein exemplarischer Fall*

Am 26. August 2018 kam es am Rande eines Stadtfests in Chemnitz zu einer gewalttätigen Auseinandersetzung. Ein 35-jähriger Mann wurde dabei getötet und zwei Freunde des Mannes verletzt. Das brutale Verbrechen, mutmaßlich begangen von einem oder zwei Männern aus Syrien und dem Irak, trat in den Social Networks noch in derselben Nacht eine Welle von Gerüchten los: Der Getötete sei angegriffen worden, weil er eine deutsche Frau vor einem sexuellen Übergriff gerettet habe, ein zweites Opfer sei nun im Krankenhaus gestorben, eine ganze „Horde" von Migranten sei über das Opfer hergefallen, die Politik wolle den Vorfall vertuschen, der „Mainstream"-Rundfunk verschweige wichtige, längst erwiesene Fakten… Ein prominenter AfD-Politiker twitterte, nun sei es „Bürgerpflicht, die todbringende Messermigration zu stoppen".[3] Neben diesen Gerüchten und Interpretationen verbreitete sich auch der Aufruf zu einem „Trauermarsch" viral, der von einer rechtsextremen Hooligan-Gruppe stammte und dem bereits am nächsten Tag rund 800 Menschen folgten – darunter militante Rechtsradikale, aber auch Bürger, die später in Gesprächen mit der Presse jede Nähe zum rechtsradikalen Milieu von sich wiesen und meinten, sie hätten nur ihre Anteilnahme und Sorge ausdrücken wollen.[4] Am Rande der Demonstration wurde jedoch der Hitlergruß gezeigt, ein jüdisches Restaurant belagert und Passanten, die für Migranten gehalten wurden, angegriffen.[5] Letzteres ging nicht nur aus Augenzeugen-

3 Vgl. https://www.tagesschau.de/faktenfinder/inland/chemnitz-geruechte-gewalt-101.html, https://www.sueddeutsche.de/politik/ausschreitungen-eine-stadt-ausser-kontrolle-1.4106465 (Stand: 24.6.2019).

4 https://projekte.sueddeutsche.de/artikel/politik/was-in-chemnitz-gschah-e866148/?reduced=true; https://www.zeit.de/2018/36/rechtsextreme-gewalt-chemnitz-regierung-mob-schock/seite-3 (Stand: 24.6.2019).

5 https://projekte.sueddeutsche.de/artikel/politik/was-in-chemnitz-gschah-e866148/?reduced=true (Stand: 24.6.2019).

berichten, sondern auch aus einer Amateur-Videoaufnahme hervor, die sich auf Twitter verbreitete und die das öffentlich-rechtliche Fernsehen zeigte.[6] Tagelang kam Chemnitz mit Demonstrationen und Gegendemonstrationen nicht zur Ruhe.

Doch die Krise blieb nicht auf Chemnitz beschränkt: Der Streit, welchen Medien, Institutionen und Auslegungen zu trauen ist, entzweite sogar zentrale politische Akteure in Deutschland: Der damalige Verfassungsschutzpräsident Hans-Georg Maaßen äußerte in einem Interview mit der Bildzeitung den Verdacht, dass die Berichterstattung des öffentlich-rechtlichen Fernsehens ein gefälschtes Video enthalten habe und dass es sich bei dem Video um eine „gezielte Falschinformation" gehandelt haben könne.[7] Heikel daran war insbesondere, dass Bundeskanzlerin Angela Merkel sich in einer Äußerung über die Vorfälle in Chemnitz auf genau dieses Video bezogen hatte.[8] Der Fälschungsverdacht von Maaßen beruhte nicht auf Erkenntnissen seiner Behörde und war nicht haltbar. Am 5. November 2018 bat Bundesinnenminister Horst Seehofer schließlich den Bundespräsidenten, Maaßen mit sofortiger Wirkung in den einstweiligen Ruhestand zu versetzen. Anlass war eine Rede Maaßens, in der er seine umstrittenen Äußerungen zu den Ausschreitungen von Chemnitz verteidigte und die laut Seehofer „inakzeptable Formulierungen" enthielt.[9]

Der Fall zeigt: Digitale Desinformation stellt demokratische Gesellschaften vor große Herausforderungen. In unklaren, krisenhaften Situationen können sich Falschinformationen viral verbreiten und dazu beitragen, dass

6 https://www.zeit.de/2018/36/rechtsextreme-gewalt-chemnitz-regierung-mob-schock/seite-3 (Stand: 24.6.2019).

7 https://www.bild.de/bild-plus/politik/inland/politik-inland/verfassungsschutz-chef-maassen-keine-information-ueber-hetzjagden-57111216,jsRedirectFrom=conversionToLogin,view=conversionToLogin.bild.html (Stand: 24.6.2019).

8 https://www.zeit.de/politik/deutschland/2018-09/chemnitz-hans-georg-maassen-hetzjagd-beweise-horst-seehofer (Stand: 24.6.2019).

9 https://www.focus.de/politik/deutschland/staatsanwaltschaft-bestaetigt-ermittlungen-nach-umstrittenem-chemnitz-video-verpruegelten-rechte-ein-deutsches-maedchen_id_9570994.html; https://www.focus.de/politik/deutschland/chronologie-der-lange-streit-um-hans-georg-maassen_id_9855929.html; https://www.zeit.de/politik/deutschland/2018-09/chemnitz-video-sachsen-hans-georg-maassen-verfassungsschutz-angriff-mob-fakten; https://www.zeit.de/politik/deutschland/2018-09/chemnitz-video-sachsen-hans-georg-maassen-verfassungsschutz-angriff-mob-fakten; https://www.zeit.de/politik/deutschland/2018-11/horst-seehofer-versetzt-hans-georg-maassen-in-einstweiligen-ruhestand; https://www.sueddeutsche.de/politik/maassen-ruhestand-seehofer-1.4197824 (Stand: 14.8.2019).

eine Stadt in eine Art Ausnahmezustand gerät. Dabei kann Desinformation für den demokratischen Prozess selbst dann schädlich sein, wenn sie „nur" in einer bestimmten Stadt, einer bestimmten Krisensituation oder einem bestimmten Milieu wirkt. Und: Die Verbreitung von Falschinformationen in Social Networks schwächt potentiell auch die faktenbasierte Berichterstattung seriöser Medien und erschwert es politischen Akteuren, Behörden, Medien und Bürgerinnen und Bürgern sich ein verlässliches Bild zu machen. Um sich dieser Herausforderung zu stellen, ist interdisziplinäre Forschung über die Grenzen klassischer Fakultäten hinweg notwendig.

B. Digitale Desinformation als Phänomen, das interdisziplinär untersucht werden muss

Mit welchen Fragen sich interdisziplinäre Forschung zum Thema digitale Desinformation beschäftigen muss, lässt sich ebenfalls anhand des beschriebenen Falls aufzeigen: Was trägt dazu bei, dass eine in Social Networks rezipierte Nachricht von Bürgern geglaubt oder sogar geteilt wird? Wie müssten Warnhinweise gestaltet sein und wie muss die Entlarvung von digitaler Desinformation (Debunking) kommuniziert werden, damit ihre Wirkung eingedämmt wird? Welche technische Unterstützung ist realisierbar, damit schnell und zugleich datenschutzkonform erkannt werden kann, dass eine Desinformations-Welle anrollt? Kann einer solchen Welle auch mit technischen Mitteln entgegengewirkt werden, ohne die freie Debatte einzuschränken? Ohne grundlegende Forschung der Medienpsychologie, die eng mit anderen Disziplinen wie Informatik und Rechtswissenschaft zusammenarbeitet, können keine qualifizierten Empfehlungen an in Krisensituationen agierende Behörden und Medien ausgesprochen werden.

Juristen wiederum müssen Antworten auf die Frage finden, wer eine Falschinformation verantwortet, die sich im Netz verbreitet. Nur der ursprüngliche Absender, der manchmal gar nicht mehr auszumachen ist? Auch Personen, die die Information weiterverbreiten und ihr dadurch erst Wirkung verleihen? Welche Verantwortung tragen Netzwerkbetreiber wie Facebook und Twitter, auch angesichts der Tatsache, dass die Betreiber die besten Möglichkeiten haben, koordinierte Desinformationskampagnen und technische Manipulationsformen wie Malicious Bots aufzuspüren? Bezüglich des Umgangs mit Verdächtigungen und mutmaßlichen Tätern hat der Deutsche Presserat für traditionelle und für Online-Medien, also Telemedien mit journalistisch-redaktionellen Inhalten, im Rahmen des Pressekodex Regeln

aufgestellt.[10] Aber wie gut funktionieren diese noch, wenn in Social Networks politische, mediale und private Akteure, die sich an dieses Regelwerk nicht gebunden sehen, zumindest lokal und in Extremsituationen hohe Reichweiten erreichen? Auch die Debatte um das vom öffentlich-rechtlichen Rundfunk gesendete Twitter-Video über die von „Trauermarsch"-Demonstranten ausgehenden Übergriffe wirft Fragen auf, die nur interdisziplinär beantwortet werden können: Wie müssen Medien Materialen und Informationen prüfen, die sie aus Social Networks übernehmen? Neue Regelungen sowohl auf gesetzlicher Ebene als auch auf der Ebene von Selbstverpflichtungen müssen gefunden werden, die die Verbreitung von Desinformation eindämmen, ohne die Verwendung wichtiger Informationen und Materialien aus Social Networks unmöglich zu machen. Notwendige Grundlage für solche Neuregelungen sind aus verschiedenen Disziplinen stammende Erkenntnisse über typische Verbreitungswege und Struktur von Desinformation, Wirkungsmacht und Verifikationsmöglichkeiten.

Auch die Journalistik muss sich mit neuen Fragen beschäftigen und kann dies nicht ohne die Unterstützung anderer Disziplinen tun: Obwohl die technischen Möglichkeiten längst sogenannte „deep fakes", also täuschend echt wirkende Videos mit gefälschten Inhalten möglich machen, werden Videos von den meisten Rezipierenden immer noch als sehr authentisch erachtet. Das stellt Medienschaffende vor Herausforderungen und macht die Zusammenarbeit mit Experten für Bild und Videoforensik notwendig.[11] Im Fall Chemnitz war das Video der gewalttätigen Ausschreitungen zwar – wie sich herausstellte – korrekt verifiziert worden, aber dass der Präsident des Bundesamts für Verfassungsschutz und hochrangige Politiker öffentlich Zweifel hegten, verstärkte den Vertrauensverlust in etablierte Medien, der in einigen gesellschaftlichen Teilmilieus besteht.[12] Wie kann diesem Vertrauensver-

10 https://www.presserat.de/fileadmin/user_upload/Downloads_Dateien/Pressekodex2017_web.pdf (Stand: 24.6.2019).

11 https://www.swr.de/swr2/leben-und-gesellschaft/it-forensiker-martin-steinebach-hat-das-strache-video-auf-echtheit-ueberprueft,article-swr-13412.html (Stand: 14.8.2019).

12 Insgesamt sollte jedoch zwischen Vertrauen in das Mediensystem als Ganzem und einzelnen Vertrauensobjekten bzw. -ebenen unterschieden werden, wie die Mainzer Langzeitstudie Medienvertrauen wiederholt für 2018 belegt, vgl. Jackob, Schultz, Jakobs et al. (2019: 211; 215-217). Während das Vertrauen in die etablierten Qualitätsmedien relativ hoch ist, steht jeder zweite Befragte Online-Informationen aus Social Networks massiv skeptisch gegenüber. Darüber hinaus stellen Schindler et al. in einer aktuellen Studie fest, dass populistische Einstellungen Medienfeindlichkeit fördern. Als Ursache sehen die Autoren eine als gering wahrgenommene Interessenvertretung

lust entgegengewirkt werden? Was bedeutet es für die Glaubwürdigkeit der traditionellen Medien, wenn die Medien sich an Sprachregeln halten, die bei einem Verdacht an sich angemessen sind, wenn sich zugleich aber in einer Stadt eine Mischung von gefälschten und tatsächlichen Details viral verbreiten, so dass viele Menschen denken, sie wüssten wesentlich mehr, als das, was in der Zeitung steht und im Rundfunk gesendet wird?

Allein aus dem „Fall Chemnitz" lassen sich also zahllose Fragen ableiten, die auf viele weitere Fälle angewendet und nur interdisziplinär beantwortet werden können. Das Beispiel zeigt, so wie sehr viele andere aktuelle Beispiele auch, dass das Phänomen der digitalen Desinformation besser untersucht werden muss – auch abseits der ohne Zweifel wichtigen Frage, wie Fake News, Malicious Bots und andere Formen der Desinformation im unmittelbaren Vorfeld von Wahlen wirken. Eine solche interdisziplinäre Untersuchung ist Gegenstand des vorliegenden Bandes.

C. *Begriffsbestimmung: Desinformation, digitale Desinformation, Malicious Bots und Fake News*

In Anlehnung an die Definition der Expertengruppe für Fake News und Desinformation der EU-Kommission wird **Desinformation hier als „falsche, ungenaue oder irreführende Informationen definiert, die erfunden, präsentiert und verbreitet werden, um Gewinne zu erzielen oder bewusst öffentlichen Schaden anzurichten"** (European Commission, 2018: 11).[13] Der Begriff der Desinformation wird also bewusst breit gefasst, denn wirkungsmächtige Desinformationskampagnen basieren nicht ausschließlich auf frei erfundenen Tatsachenbehauptungen, sondern auf einer Mischung von wahren Begebenheiten, noch nicht bewiesenen (aber auch noch nicht

durch Medien sowie die Nutzung alternativer Medien. Zudem seien Rezipienten mit derart entstandenen medienfeindlichen Einstellungen durch eine höhere politische Aktivität sowie häufigere Meinungsäußerungen in den Medien sichtbarer als Menschen, die Medien gegenüber weniger feindlich gegenüberstehen, vgl. Schindler, Fortkord, Posthumus et al. (2018: 293-296). Zum Medienvertrauen und der Wahrscheinlichkeit, Desinformation zu glauben, hat auch die Stiftung Neue Verantwortung eine Befragung durchgeführt, die zeigt, dass AfD-Wähler klassischen Medien weniger trauen und eher gewillt sind, Fake News zu glauben, vgl. Sängerlaub, Meier & Rühl (2018: 87-88).

13 Die hier zitierte Übersetzung ins Deutsche wurde aus der zum Bericht gehörenden deutschsprachigen Pressemitteilung übernommen: http://europa.eu/rapid/press-release_IP-18-1746_de.htm (Stand: 12.6.2019).

widerlegten) Gerüchten und Verdächtigungen und hinzuerfundenen Details (European Commission 2018: 10). Die in diesem Band behandelten Desinformationen umfassen sowohl rechtmäßige als auch rechtswidrige Inhalte.

Desinformationen greifen dabei, wie sich auch am Beispiel von Chemnitz zeigt, oftmals reale Begebenheiten auf und bringen vorhandene Stimmungen zum Ausdruck, um dann Verallgemeinerungen abzuleiten, die nicht folgerichtig sind und Emotionen auf Ziele zu lenken, deren Involvierung auf unzulässigen Verallgemeinerungen und Schlussfolgerungen beruht. **Desinformation zielt darauf ab, zu manipulieren, also einen falschen Eindruck zu erwecken.** Dies kann durch die Beimischung erfundener Details geschehen, durch den Einsatz von optischen Falschinformationen, wie manipulierten oder aus einem anderen Kontext stammenden Bildern und Videos, aber auch dadurch, dass Malicious Bots oder massenhaft hinzugekaufte Follower ein Stimmungsbild vermitteln, das nicht den realen Gegebenheiten entspricht. Im vorliegenden Band legen wir den Schwerpunkt auf Desinformationskampagnen, die politische und gesellschaftliche Themen betreffen.[14]

Die bis hier geschilderten Eigenschaften von Desinformation sind dabei weder neu noch auf den digitalen Raum beschränkt: Bereits im Kalten Krieg wurden durch Desinformationskampagnen in der Bevölkerung bestehende Ängste und Stimmungen gezielt aufgegriffen und in Artikeln wissenschaftlich erwiesene Fakten mit erfundenen Informationen angereichert. Ein Beispiel ist die Desinformationskampagne, die in den 80er Jahren Aids als Erfindung der US-Armee darstellte (u.a. Boghardt, 2009). Historisch breit angelegte Studien zeigen, dass es bereits vor der Erfindung des Internets Phasen mit besonders hoher Desinformationsdichte gab (vgl. Butter 2014) und neu aufkommende Medien immer wieder besonders stark für die Verbreitung manipulierender Mythen genutzt wurden (Blume, 2019).

Trotz dieser interessanten historischen Vergleichsfälle beschränkt sich die in diesem Band dargestellte Forschung explizit auf Desinformation, die im Internet verbreitet wird.[15] Was zeichnet digitale Desinformation also aus und warum lohnt es, diese gesondert zu untersuchen? Wichtig ist es

14 Die von der EU-Kommission bestellte Expertengruppe beschäftigt sich explizit auch mit Themen aus Bereichen wie Gesundheit, Wissenschaft, Bildung und Finanzen, vgl. European Commission (2018: 10; 24). Zu Falschinformationen im Gesundheitsbereich s. z. B. Feldwisch-Drentrup & Kuhrt (2019).

15 Dabei sind wir uns bewusst, dass digitaler Desinformation oftmals intrapersonale Desinformation vorausgeht oder nachfolgt: Im persönlichen Gespräch aufgeschnappte Gerüchte verbreiten sich digital weiter und online rezipierte Desinformation wird in Gesprächen weitergegeben, vgl. Zimmermann & Kohring (2018: 537).

zunächst festzustellen, dass die Unterschiede zwischen Desinformation, die sich online verbreitet, und solcher, die sich offline verbreitet, alle gradueller und nicht absoluter Natur sind. Dennoch stellt das Internet im Allgemeinen und stellen Social Networks im Besonderen ein Umfeld dar, das sich in Bezug auf die Verbreitung von Desinformation durch eine Reihe von Besonderheiten auszeichnet:

Digitale Desinformation kann sich sehr schnell (*Tempo*; Huxford, 2007; Kümpel, 2018; Vosoughi et al., 2018) verbreiten. Dies hat zur Folge, dass sie gerade in Krisensituationen über die Nutzung mobiler Endgeräte bei vielen Menschen im jeweils betroffenen Gebiet oder der betroffenen Stadt ankommen und somit auch auf die konkrete Krisensituation zurückwirken kann (*Reziprozität*; Fathi et al., 2019). Darüber hinaus sind die Verbreitungswege für digitale Desinformationskampagnen kostengünstig (*Kosten*; Gerhards & Schäfer, 2007; Haim, 2019), jedermann zugänglich und ohne besonderen Verschleierungsaufwand anonym nutzbar (*Anonymität*; Brodnig, 2013; Brunst, 2009). Für die massenhafte Verbreitung von digitaler Desinformation sind zwar in der Regel Akteure und Kanäle mit vorab hohen Follower-Zahlen notwendig, in Einzelfällen und gerade in Krisensituationen können aber Dynamiken entstehen, durch die sich Desinformationen, die ursprünglich von unbedeutenden Quellen stammen, massenhaft verbreiten (*potentielle Massenwirksamkeit*; Kümpel, 2018; Silverman, 2016). Digitale Desinformationskampagnen können zudem durch die Nutzung bestimmter Gruppen und Plattformen und die Möglichkeiten des Microtargeting sehr spezifisch auf Teil-Zielgruppen zugeschnitten werden (*Passgenauigkeit*; Haim, 2019; Mitchelstein & Boczkowski, 2009). Hinzu kommt, dass digital verbreitete Desinformation oft lange unbemerkt bleibt.[16] Dies kann zum einen daran liegen, dass gerade lokale Behörden digital weniger präsent sind als im öffentlichen Raum. Dass digitale Kampagnen oft nicht richtig eingeschätzt werden, kann aber auch darauf zurückgeführt werden, dass online die Bedeutung von großen, aber dennoch geschlossenen digitalen Gruppen wächst und zunehmend auch Messenger-Dienste wie beispielsweise WhatsApp zur Verbreitung von Desinformation genutzt werden (*Unsichtbarkeit/Dark Social*; Kümpel, 2018)[17]. **Das Internet im Allgemeinen und Social Networks im Besonderen bieten also ein Umfeld für die Verbreitung von Desinformation, das**

16 So wurde beispielsweise die Reichweite und Wirkungsmacht des Aufrufs zum Trauermarsch von der Polizei in Chemnitz unterschätzt, vgl. https://www.zeit.de/politik/deutschland/2018-08/ausschreitungen-in-chemnitz-sachsen-polizei-versagen (Stand: 24.6.2019).

17 Zur wachsenden Bedeutung von geschlossenen Gruppen siehe z. B. Swart et al., 2018.

sich durch ein hohes Tempo, Reziprozität, niedrige Kosten, Anonymität, Massenverbreitung, Passgenauigkeit und Unsichtbarkeit auszeichnet.

Digitale Desinformation wird im vorliegenden Band als übergeordnetes Phänomen verstanden, das verschiedene, jedoch oftmals zusammenwirkende Desinformationsphänomene, -methoden und -mechanismen beinhaltet: Dazu gehören Fake News, also pseudojournalistische Inhalte, die über entsprechende Webseiten und Social Networks verbreitet werden, aber auch nicht-journalistische Desinformation, die ebenfalls auf die Manipulation der öffentlichen Meinung abzielt, wie beispielsweise nicht authentische Augenzeugenberichte und Vorort-Videos, gefälschte Behördenschreiben etc.[18] Auch der Einsatz von Malicious Bots und Fake-Accounts wird, soweit sie darauf abzielen, das öffentliche Meinungsbild zu manipulieren, von uns dem Phänomen der digitalen Desinformation zugerechnet.

In Anlehnung an Ferrara et al. (2016) werden Social Bots definiert als Profile in Social Networks, die vollständig oder teilweise von Algorithmen gesteuert werden, automatische Inhalte generieren, mit weiteren Nutzenden (Menschen und Maschinen) interagieren können und sich dabei als echte menschliche Nutzende ausgeben oder diese gut nachahmen. **Social Bots werden häufig zur Verbreitung von Fake News, gefälschten Augenzeugenberichten und User-Kommentaren eingesetzt und gelten somit als wichtige Instrumente in Desinformations-Kampagnen. In solchen Fällen werden sie als Malicious Bots (wörtlich übersetzt bösartige Bots) bezeichnet**. Algorithmisch gesteuerte Social-Network-Konten können aber auch harmlos oder gar hilfreich sein, z. B. wenn monotone Tätigkeiten wie einfache Gespräche mit Menschen ausgeführt werden müssen. Derartige Konten gelten allgemein als gutartige (benign) Social Bots (Ferrara et al., 2016).

Der Begriff Fake News wiederum bedarf ebenfalls einer kurzen Diskussion, schon aufgrund der Tatsache, dass er regelmäßig wissenschaftliche Kontroversen entfacht und verschiedene Deutungsvarianten bestehen. Manche Wissenschaftler lehnen den Begriff insgesamt ab, weil er inzwischen oftmals auch von Politikern wie dem US-Präsidenten Donald Trump benutzt wird, um faktenbasierte Nachrichten in Frage zu stellen, die über seriöse Medien verbreitet werden und ihnen missfallen (European Commission, 2018: 5; Tandoc Jr et al., 2018; Vosoughi et al., 2018; Zimmermann & Kohring, 2018). In diesem Band wird neben dem übergeordneten Begriff der Desinformation auch der Begriff Fake News benutzt, trotz der teilweise berechtigten

18 Solche Desinformation wird oftmals im Rahmen von Fake News aufgegriffen, verbreitet sich teilweise auch, ohne im pseudojournalistischen Stil aufbereitet zu werden.

Einwände. Zum einen weil er als analytische Kategorie auf präzise und allgemeinverständliche Art und Weise beschreibt, was hier untersucht wird, zum anderen weil die Autorinnen angelehnt an Lazer et al. (2018: 1094) davon ausgehen, dass Wissenschaftler den Begriff nicht dem populistischen Re-Framing überlassen sollten, sondern die ursprüngliche Bedeutung des Begriffs öffentlich re-etablieren sollten.

Im vorliegenden Band werden **Fake News als online verbreitete Informationen definiert, die journalistische Nachrichteninhalte nachahmen, indem sie journalistische Routinen der Nachrichtenpräsentation und -auswahl anwenden, bei deren Produktion zugleich aber journalistische Rechercheroutinen systematisch missachtet werden und deshalb falsche Tatsachenbehauptungen enthalten** (ähnlich Lazer et al., 2018: 1094). Die falschen Behauptungen können dabei durch standardisierte journalistische Fact-Checking-Routinen entlarvt werden. Zugleich ähnelt ihr Erscheinungsbild den Produkten journalistischer Onlinemedien, die diese Regeln beachten. Sie täuschen also vor, vertrauenswürdige Informationsgrundlagen zu sein.

D. Aufbau des Bandes und interdisziplinäres Forschungsanliegen

Kommunikationswissenschaftlerinnen, Informatiker, Medienpsychologinnen und Juristen setzen sich im vorliegenden Band mit unterschiedlichen Aspekten der digitalen Desinformation auseinander. Dabei wird einerseits ein Überblick über den Forschungsstand zum Thema Desinformation in der jeweiligen Disziplin gegeben und werden aktuelle Forschungsergebnisse der jeweiligen Autorinnen dargestellt. Andererseits wird aber über zahlreiche Querverweise sowie das gemeinsame Schlusskapitel abgebildet, dass die umfassende Erforschung des Phänomens der digitalen Desinformation nur im stetigen, interdisziplinären Austausch erfolgen kann.

In Kapitel 2, das dieser Einleitung folgt, wird die zentrale Rolle von Fake News für Desinformationskampagnen im deutschsprachigen Internet aus Sicht der Journalistik beschrieben. Hierbei geht es – anders als in den meisten anderen auf den deutschen Sprachraum bezogenen Forschungsprojekten zu Fake News –, nicht ausschließlich um die Verbreitung von Pseudojournalismus im unmittelbaren Umfeld von Wahlen (Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018). Vielmehr wird ganz grundsätzlich erfasst, welche strukturellen, sprachlichen, argumentativen und thematischen Muster Fake News im deutschsprachigen Raum aufweisen. Erstellt und untersucht wurde hierfür ein Sample von fast 500 Fake News, die zwischen

Dezember 2015 und März 2018 im deutschsprachigen Internet veröffentlicht wurden. Das große Sample und der lange Zeitraum ermöglichen es erstmals nachzuvollziehen, wie Fake-News-basierte digitale Desinformation im deutschsprachigen Raum zur Entstehung von Narrativen beiträgt, die dann vor Wahlen, aber auch in Krisensituationen, wie während der Ereignisse in Chemnitz, aktiviert werden können.

In Kapitel 3 wird aus medienpsychologischer Perspektive untersucht, wann Desinformation von Rezipienten als wahr angenommen wird. Hierzu werden Merkmale der Nachricht, der Quelle und der Rezipienten und Rezipientinnen in Augenschein genommen. Aufbauend auf den Erkenntnissen über sprachliche und strukturelle Eigenschaften von Fake News wird aufgezeigt, welche der Merkmale, die als charakteristisch für Fake-News-Inhalte beschrieben wurden (vgl. Kap. 2), die Wirkung von Fake News verstärken. Dabei werden Merkmale wie reißerische Formulierungen, Inkonsistenzen, Subjektivität, unglaubwürdige Quelle oder manipuliertes Bild variiert und die jeweilige Wirkung untersucht. Darüber hinaus wird untersucht, ob positive oder negative Desinformationen über Politikerinnen und Politiker eher geglaubt werden und wie Meinungsbildung in einem Umfeld, in dem Malicious Bots aktiv sind, manipuliert werden kann. Weil die Löschung bedenklicher Inhalte, deren Gesetzeswidrigkeit aber noch nicht abschließend festgestellt wurde, aus rechtsstaatlicher Sicht und im Sinne der freien Meinungsäußerung hochproblematisch ist (vgl. Kap. 5), setzt sich die Medienpsychologie außerdem mit der Frage auseinander, wie Warnhinweise formuliert und dargestellt werden müssen, damit sie von den Rezipierenden wahrgenommen werden.

Kapitel 4 lotet aus Sicht der Informatik Möglichkeiten zur automatisierten Erkennung von Desinformation im Internet aus. Die Medientypen Text, Bild und Video werden dabei gesondert betrachtet, denn jeder Medientyp weist eigene Manipulationstypen und Erkennungsmethoden auf. Das im journalistischen Teilprojekt generierte Sample von Fake-News-Texten wird beispielsweise eingesetzt, um Algorithmen darauf zu trainieren, verdächtige Texte als solche zu markieren. Auf dieser Grundlage kann dann Netzwerkbetreibern oder auch Nutzende eine weitere Überprüfung empfohlen werden. Um manipulierte Bilder, Videos und Audios zu erkennen, wird wiederum Multimedia-Forensik angewandt. Methoden aus der Bildforensik können dabei nicht nur genutzt werden, um gefälschte Augenzeugenvideos und andere nicht-journalistische digitale Desinformation zu enttarnen, sondern sie sind auch maßgeblich, um seriösen Journalismus und Fake News dauerhaft unterscheidbar zu halten: Wenn Fake News, wie in diesem Band definiert, sich dadurch auszeichnen, dass ihre Produzenten journalistische Fact-Checking-

Regeln systematisch missachten, muss sich im Umkehrschluss seriöser Journalismus dadurch auszeichnen, dass er routinemäßig zeitgemäße und dem neusten technischen Stand entsprechende Maßnahmen zur Quellenüberprüfung vornimmt – beispielsweise in Sachen Videoforensik.[19] Und – das zeigt das Video aus Chemnitz und der Fall Maaßen – auch die Sicherheitsbehörden sollten solche Überprüfungen unbedingt durchführen, bevor sie beispielsweise eine von Medienschaffenden verwendete Quelle öffentlich anzweifeln. Ein weiterer Aspekt digitaler Desinformation, der in Kapitel vier aufgearbeitet wird, ist die Erkennung von Malicious Bots. Weil die von außerhalb der Netzwerke anwendbaren Erkennungsmethoden sich in der Praxis noch als recht unzuverlässig erweisen, müssen hier neue Wege gegangen werden, falls Maßnahmen von außen zur Eindämmung von Malicious Bots umgesetzt werden sollen (vgl. Kap. 5 und Kap. 3). Der Schwerpunkt von Kapitel 4 liegt auf der Erkennung von digitaler Desinformation durch Technik. Hier wird skizziert, wie die Authentizität und Integrität von Zitaten technisch nachgewiesen werden kann. Diese Verfahren könnten auch hilfreich für die Qualitätssicherung in Redaktionen sein.

Weil juristische Maßnahmen, auch auf der Ebene von Selbstverpflichtungen, für die Bekämpfung von digitaler Desinformation eine zentrale Rolle spielen und zugleich nur erfolgreich sein können, wenn sie auf präzisen Kenntnissen über Struktur, Wirkung und technische Erkennbarkeit von Desinformation basieren, schließt das Kapitel über rechtliche Bekämpfungsmöglichkeiten den Reigen der disziplinären Einzelbetrachtungen in diesem Band ab (Kap. 5). Ausgangspunkt ist, dass in Deutschland kein generelles Gesetz existiert, das die Herstellung und Verbreitung von Falschnachrichten untersagt – vielmehr müssen je nach Tatbestand ganz unterschiedliche Rechtsnormen angewandt werden, wobei diese Rechtsnormen teilweise noch weiter präzisiert werden müssen, um digitaler Desinformation sinnvoll entgegenwirken zu können. Stets berücksichtigt werden muss dabei, dass es sich hier um einen demokratisch hochsensiblen Bereich handelt und die Kommunikationsgrundrechte vollständig gewahrt werden müssen. Das Recht auf

19 Ein gelungenes Beispiel für eine solche Zusammenarbeit zwischen Wissenschaftlern und Medienschaffenden in Sachen Videoforensik stellt die Verifikation des Videos dar, das den damaligen FPÖ-Parteichef Strache auf Ibiza zeigt. Ein derart brisantes Video, das zudem mit unklaren Motiven und von Unbekannten produziert wurde, kann von seriösen Medien nur veröffentlicht werden, wenn eine hohe gesellschaftliche Relevanz vorliegt und zudem jeder Fälschungsverdacht ausgeräumt wurde. Dazu: https://www.spiegel.de/video/strache-video-oesterreich-gutachter-prueften-der-aufnahmen-video-99027197.html (Stand: 13.6.2019).

freie Meinungsäußerung schützt dabei auch sogenannte Mischäußerungen, in denen Meinung und (falsche) Tatsachenbehauptung untrennbar miteinander verbunden sind, was bei Desinformationen oftmals der Fall ist (vgl. Kap. 2.2). Zugleich spielen Informationen und ihre inhaltliche Qualität eine wichtige Rolle für den öffentlichen Meinungsbildungsprozess und nicht zuletzt für die politische Öffentlichkeit, da sie die Bürgerinnen und Bürger befähigen sollen, sachkundig an öffentlichen Diskussionen und Wahlen zu partizipieren.[20] Die Autoren gehen daher der Frage nach, ob und inwieweit von Desinformationen betroffene Personen und die Allgemeinheit vor der Verbreitung von Unwahrheiten sowie Manipulationsformen wie Malicious Bots geschützt werden können und müssen. Ein Fokus liegt auf der besonderen rechtlichen Verantwortung großer Social Networks, die einerseits gesetzliche Lösch- und Sperrverpflichtungen nach Meldung von rechtswidrigen, insbesondere strafbaren Inhalten zu erfüllen haben und die andererseits unter Verweis auf ihre AGB (Community-Richtlinien, Gemeinschaftsstandards) nutzergenerierte Inhalte regulieren. In dieser Hinsicht stellen sich bisher ungeklärte Fragen im Hinblick auf die Reichweite ihrer Befugnisse, Inhalte zu entfernen oder in der Sichtbarkeit einzuschränken und Konten von Nutzenden zu sperren und zu entfernen. Als problematisch erweist sich zudem, dass Selbstregulierungsmaßnahmen wie die Regeln des Deutschen Presserates von vielen neuen Onlinemedien nicht anerkannt werden. Vor diesem Hintergrund könnten neue Regelungen notwendig werden. So diskutieren die Autoren, ob besonders schwere Verstöße gegen die Sorgfaltspflichten für journalistisch-redaktionelle Onlinemedien durch eine Erweiterung der Befugnisse der Landesmedienanstalten geahndet werden könnten.

Leserinnen und Leser dieses Bandes werden also Antworten auf grundlegende Fragen zum Thema der digitalen Desinformation finden: Was macht die Desinformation im deutschsprachigen Internet aus (Kapitel 2)? Wie wirkt Desinformation (Kapitel 3)? Wie kann sie mithilfe technischer Mittel erkannt werden (Kapitel 4)? Was kann und könnte mit (selbst-)regulatorischen und rechtlichen Maßnahmen gegen Desinformation getan werden (Kapitel 5)?

Aus den größtenteils explorativ angelegten disziplinären Studien leiten sich jedoch nicht nur Erkenntnisse über digitale Desinformation ab, sondern auch ein weiterer interdisziplinärer Forschungsbedarf, der in Kapitel 6 dargestellt wird. Darüber hinaus ergeben sich aus den Erkenntnissen aller Teilbereiche Handlungsempfehlungen für Gesetzgeber, Presserat, Medien-

20 Holznagel, NordÖR 2011, 205 (209).

schaffende, Betreiber von Social Networks und nicht zuletzt Mediennutzende, die ebenfalls im abschließenden Kapitel dargestellt werden.

Das Phänomen der digitalen Desinformation sollte nicht überbewertet werden und die Panik, die regelmäßig im unmittelbaren Vorfeld von Wahlen ausbricht, war soweit stets unangebracht – dennoch zeigt dieser Band: Digitale Desinformation stellt in Deutschland ein ernstzunehmendes Problem dar. Sie trägt zur Entstehung von Stimmungen in Teilbereichen der Gesellschaft bei, die im Einzelfall zu Gewalt führen können und in Krisensituationen aktiviert werden können. Zudem zeigt dieser Band: Digitale Desinformation kann nur eingedämmt werden, wenn in Politik, Behörden, Medien und bei den Bürgerinnen und Bürgern ein Bewusstsein für die Mechanismen und Wirkungsweisen von Desinformation besteht und der einschlägige gesellschaftliche Diskurs fortgesetzt wird.

Literaturverzeichnis zu Kapitel 1

Blume, M. (2019). Warum der Antisemitismus uns alle bedroht. Wie neue Medien alte Verschwörungstheorien befeuern. Ostfildern: Patmos Verlag.

Boghardt, T. (2009). Soviet Bloc Intelligence and its AIDS Disinformation Campaign. Studies in Intelligence, 53 (5), 1-24.

Brodnig, I. (2013). Der unsichtbare Mensch. Wie die Anonymität im Internet unsere Gesellschaft verändert. Wien: Czernin.

Brunst, P. W. (2009). Anonymität im Internet - rechtliche und tatsächliche Rahmenbedingungen : Zum Spannungsfeld zwischen einem Recht auf Anonymität bei der elektronischen Kommunikation und den Möglichkeiten zur Identifizierung und Strafverfolgung. Duncker & Humblot.

Butter, M. (2014). Plots, Designs, and Schemes: American Conspiracy Theories from the Puritans to the Present (linguae & litterae, Vol. 33). Berlin: De Gruyter. Retrieved http://www.degruyter.com/search?f_0=isbnissn&q_0=9783110346930&searchTitles=true.

Egelhofer, J. L. & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. Annals of the International Communication Association, 43(2), 97-116.

European Commission. (2018). A multi-dimensional approach to disinformation - Report of the independent High Level Group on fake news and online disinformation. Zugriff am 18.4.2018. Verfügbar unter https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

Fathi, R., Kleinebrahn, A., Schulte, Y. & Martini, S. (2019, 06. Februar). Desinformation in der Lage oder - Die Suche nach dem Koch der Gerüchteküche. Verfügbar unter https://crisis-prevention.de/innere-sicherheit/desinformationen-in-der-lage-oder-die-suche-nach-dem-koch-der-geruechtekueche Zugriff am 19.6.2019.

Feldwisch-Drentrup, H. & Kuhrt, N. (2019). Schlechte und gefährliche Gesundheitsinformationen. Wie sie erkannt und Patienten besser geschüzt werden können. (Bertelsmann Stiftung, Hrsg.). Zugriff am 4.9.2019. Verfügbar unter https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/schlechte-und-gefaehrliche-gesundheitsinformationen/

Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

Gerhards, J. & Schäfer, M. S. (2007). Demokratische Internet-Öffentlichkeit? Ein Vergleich der öffentlichen Kommunikation im Internet und in den Printmedien am Beispiel der Humangenomforschung. Publizistik, 52(2), 210-228.

Haim, M. (2019). Die Orientierung von Online-Journalismus an seinen Publika. Anforderung, Antizipation, Anspruch.

Huxford, J. (2007). The proximity paradox. Journalism: Theory, Practice & Criticism, 8(6), 657-674.

Jackob, N., Schultz, T., Jakobs, I., Ziegele, M., Quiring, O. & Schemer, C. (2019). Medienvertrauen im Zeitalter der Polarisierung. Media Perspektiven – Onli-

ne, 49(5), 210-220. Zugriff am 19.6.2019. Verfügbar unter https://www.ard-werbung.de/fileadmin/user_upload/media-perspektiven/pdf/2019/0519_Jackob_Schultz_Jakobs_Ziegele_Quiring_Schemer_2019-06-12.pdf

Kümpel, A. S. (2018). Nachrichtenrezeption auf Facebook. Wiesbaden: Springer Fachmedien Wiesbaden.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B. et al. (2018). The Science of Fake News. Science (New York, N.Y.), 359(6380), 1094-1096.

Marchal, N., Kollanyi, B., Neudert, L.-M. & Howard, P. N. (2019, 21. Mai). Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook. (Project on Computational Propaganda, Hrsg.) (Data Memo 2019.3). Oxford, UK. Zugriff am 4.6.2019. Verfügbar unter https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp_GermanElections_Sep2017v5.pdf

Mitchelstein, E. & Boczkowski, P. J. (2009). Between tradition and change. Journalism 10(5), 562-586.

Neudert, L.-M., Kollanyi, B. & Howard, P. N. (2017, 19. Juli). Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter? (Project on Computational Propaganda, Hrsg.) (Data Memo 2017.7). Oxford, UK. Zugriff am 4.6.2019. Verfügbar unter https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp_GermanElections_Sep2017v5.pdf

Sängerlaub, A., Meier, M. & Rühl, W.-D. (2018). Fakten statt Fakes: Das Phänomen "Fake News". Verursacher, Verbreitungswege und Wirkungen von Fake News im Bundestagswahlkampf 2017 (Stiftung Neue Verantwortung, Hrsg.) (Abschlussbericht Projekt "Measuring Fake News"). Berlin. Zugriff am 6.6.2018. Verfügbar unter https://www.stiftung-nv.de/sites/default/files/snv_fakten_statt_fakes.pdf

Schindler, J., Fortkord, C., Posthumus, L., Obermaier, M., Reinemann, C., Nayla & Fawzi (2018). Woher kommt und wozu führt Medienfeindlichkeit? Zum Zusammenhang von populistischen Einstellungen, Medienfeindlichkeit, negativen Emotionen und Partizipation. M&K Medien & Kommunikationswissenschaft, 66(3), 283-301.

Silverman, C. (2016, 16. November). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. Zugriff am 28.2.2017. Verfügbar unter https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.haGzOlwPjA#.btnP9wrD2O

Swart, J., Peters, C., & Broersma, M. (2018). Shedding light on the dark social: The connective role of news and journalism in social media communities. *new media & society*, 20(11), 4329-4345.

Tandoc Jr, E. C., Lim, Z. W. & Ling, R. (2018). Defining “Fake News”. Digital Journalism, 6(2), 137-153.

Vosoughi, S., Roy, D. & Aral, S. The Spread of True and False News Online (2018). Science, 359(6380), 1146-1151.

Wardle, C. (2017). *Fake News. Es ist kompliziert,* First Draft. Zugriff am 26.9.2017. Verfügbar unter https://de.firstdraftnews.org/fake-news-es-ist-kompliziert/

Zimmermann, F. & Kohring, M. (2018). „Fake News“ als aktuelle Desinformation. Systematische Bestimmung eines heterogenen Begriffs. M&K Medien & Kommunikationswis-

senschaft – Online, 66(4), 526-541. Verfügbar unter https://www.nomos-elibrary.de/10.5771/1615-634X-2018-4-526.pdf

Kapitel 2: Jenseits der Fakten: Deutschsprachige Fake News aus Sicht der Journalistik

Autoren:

Prof. Dr. Katarina Bader
Dr. Carolin Jansen
Prof. Dr. Lars Rinsdorf

A. Fake News – ein Geschäft mit der Angst?

Gleichgültig, ob es sich um eine wissenschaftliche Debatte handelt, ein mit journalistischen Praktikern besetztes Podium oder ein Stammtischgespräch: Seit 2016 fällt in Diskussionen, in denen es um die aktuelle Rolle von Onlinemedien und Social Networks im gesellschaftlichen Diskurs geht, fast zwangsläufig der Begriff „Fake News". Im Internet und über Social Networks verbreitete Falschmeldungen, die in Sachen Themenwahl, Aufbereitung und Darstellungsformen echten Journalismus imitieren, aber journalistische Recherchestandard systematisch missachten und falsche Tatsachenbehauptungen verbreiten, werden weltweit als relevantes Problem wahrgenommen. Sowohl in hochentwickelten Staaten wie den USA und Deutschland als auch in Entwicklungsländern und Schwellenländern wie beispielsweise Indien wird derartigem Pseudo-Journalismus das Potenzial zugeschrieben, das politische Klima mitzuprägen und demokratische Prozesse zu stören, insbesondere vor Wahlen.[1]

Im folgenden Kapitel wird es darum gehen, kommunikationswissenschaftliche Erkenntnisse über Fake News im deutschsprachigen Raum systematisch zusammenzutragen und eigene Forschungsergebnisse im Kontext bereits bestehender Fake-News-Forschung vorzustellen.

Einen exemplarischen Einblick in das deutsche Fake-News-Milieu gab 2018 ein Prozess in Berlin: Mit Mario R. stand dort ein langjähriger Aktivist des rechtspopulistischen Milieus vor Gericht. Er gestand, Inhaber eines On-

1 Für Indien: https://faktenfinder.tagesschau.de/ausland/indien-wahl-fake-news-101.html; https://comprop.oii.ox.ac.uk/research/india-election-memo/ (Stand: 24.6.2019) bzw. Mahatpatra & Plagemann (2019: 6).

lineshops zu sein, der unter dem bizarren Namen „Migrantenschreck" von Ungarn aus illegal Waffen nach Deutschland verkaufte.[2] Das Startkapital für diesen Waffenhandel, auch das räumte R. vor Gericht ein, stamme im Wesentlichen aus Provisionen, die er für den Abo- und Buchverkauf von rechtsgerichteten Medien wie Compact und dem Kopp-Verlag erhalten habe.[3] Presseberichten zufolge vermutet die Staatsanwaltschaft, dass R. nicht nur mit Büchern, Abos und Waffen handelte: R. wird darüber hinaus auch verdächtigt, selbst der Betreiber einer Seite namens „anonymousnews.ru" gewesen zu sein.[4] Auf dieser wurde regelmäßig Werbung für den Onlineshop „Migrantenschreck" angezeigt neben mit falschen Tatsachenbehauptungen angereicherten Meldungen, die ein Angstnarrativ inszenieren und sich gut eignen, um zum Waffenkauf zu ermuntern: Die Partei „Die Grünen" wolle in Bayern die Gefängnisse abschaffen, wurde da vermeldet, ein Pakistani in Hamburg habe ein Kind geschächtet, in Essen tobe der Sex-Dschihad und die staatliche Struktur in ganz Deutschland löse sich auf.[5] Noch ist nicht gerichtlich geklärt (Stand 24.6.2019), ob Mario R. derartige Meldungen selbst verfasst hat oder „anonymousnews.ru" nur als Werbeumfeld für seinen

2 https://www.sueddeutsche.de/panorama/waffenhandel-migrantenschreck-haftstrafe-1.4258587 (Stand: 24.6.2019).

3 https://www.sueddeutsche.de/politik/compact-und-kopp-rechte-verlage-zahlten-wohl-geld-an-betreiber-von-hetzseiten-1.4230649 (Stand: 24.6.2019).

4 Quelle: https://www.taz.de/Urteil-gegen-rechten-Waffenhaendler/!5556852/ (Stand: 24.6.2019). Dass annonymousnews.ru kein randständiges Portal ist, weisen Riedlinger und von Detten (2018) in einer Netzwerkanalyse nach, die zeigt, dass annoynmousnews.ru zu den Seiten gehört, die von den anderen Fake-News-Seiten besonders stark verlinkt werden. Allerdings überwarf R. sich bereits vor seiner Verhaftung mit einigen anderen Akteuren der Fake-News-Szene. Dass dieser Konflikt im Netz öffentlich ausgetragen wurde, ermöglicht interessante Einblicke in Produktionsweise und Geschäftsmodelle einiger deutscher Fake-News-Akteure. Siehe dazu: https://www.compact-online.de/83127-2/ (Stand: 24.6.2019); https://juergenfritz.com/2018/03/06/anonymousnews-ru/ (Stand: 24.6.2019).

5 *Gefängnisse* (Meldung vom 13.10.2017): https://www.anonymousnews.ru/2017/10/13/gefaengnisse-abschaffen-gruene-wollen-90-prozent-der-haeftlinge-in-offenen-vollzug-entlassen/ (Stand: 24.6.2019). *Kind* (Meldung vom 2.10.2017): https://www.anonymousnews.ru/2017/10/24/hamburg-2-jaehrigen-maedchen-kehle-durchgeschnitten-taeter-ist-fluechtling-aus-pakistan/ (Stand: 24.6.2019). *Sex-Dschihad* (Meldung vom 1.11.2017): https://www.anonymousnews.ru/2017/11/01/sex-dschihad-in-essen-200-fluechtlinge-stuermen-halloween-party-medien-schweigen/ (Stand: 24.6.2019). *Innere Ordnung* (Meldung vom 15.10.2017): https://www.anonymousnews.ru/2017/10/15/innere-ordnung-erodiert-hamburger-polizei-ist-am-ende-der-staat-beginnt-sich-aufzuloesen/ (Stand: 24.6.2019).

Waffenhandel nutzte (siehe Abb. 2.1). Dennoch zeigt der Fall: Im deutschsprachigen Internet ermöglichen Fake News perfide Geschäftsmodelle.

Abbildung 2.1: Werbung für Waffen und den Kopp-Verlag auf anonymousnews.ru

Berichte über kriminelle Flüchtlinge, Werbung für rechtspopulistische Verlage und für Waffen auf anonymousnews.ru Quelle: https://web.archive.org/web/20160705182512/http://www.anonymousnews.ru/2016/07/02/faktencheck-sind-fluechtlinge-wirklich-krimineller-als-deutsche (Stand: 24.6.2019)

Von der im folgenden dargestellten Analyse abgesehen existieren für den deutschen Sprachraum bisher nur einzelne größere empirische Studien zum Thema Fake News (Brinkschulte et al., 2019; Humprecht, 2018; Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018). Um die hier vorgestellten Erkenntnisse einzuordnen, wird deshalb auch immer wieder auf empirische Studien verwiesen, die im angloamerikanischen Raum durchgeführt wurden, für den bereits mehr empirisch abgesicherte Erkenntnisse über Fake News vorliegen (beispielsweise Allcott & Gentzkow, 2017; Benkler et al., 2018; Vosoughi et al., 2018).[6] Dabei müssen aber einige grundle-

6 Einen guten Überblick bietet Dias (2017) unter https://firstdraftnews.org/seven-studies-mis-disinformation-2017/ (Stand: 24.6.2019).

gende Unterschiede zwischen der deutschen und der US-amerikanischen Fake-News-Szene berücksichtigt werden:

In den USA ist das Potenzial, über gefälschte Meldungen Klicks und Werbeeinnahmen zu generieren, höher. Dies liegt zunächst an Hintergrundfaktoren: Die englischsprachige Zielgruppe ist per se größer und Social Networks stellen für einen großen Bevölkerungsanteil eine wichtige Nachrichten-Quelle dar (Shearer & Matsa, 2018). Darüber hinaus existieren im stark polarisierten US-amerikanischen Mediensystem Medien mit hoher Reichweite, die, wie beispielsweise FOX-News, immer wieder Falschmeldungen aufgreifen und so zu einer massenhaften Verbreitung beitragen (Benkler et al., 2018: 75). Beides führte dazu, dass dort vor der letzten Präsidentschaftswahl eine große Zahl von Fake News zirkulierte, die teilweise sehr hohe Reichweiten erreichten (Allcott & Gentzkow, 2017; Guess et al., 2018; Guess et al., 2019; Hindman & Barash, 2018; Marwick & Lewis, 2017; Nelson & Taneja, 2018; Wardle & Derekhshan, 2017).

Wer hingegen für den deutschsprachigen Raum Fake News produziert, zielt meist auf ein spezifisches rechtspopulistisches Milieu ab. Dieses Milieu ist zu groß und entwickelt sich zu dynamisch, als dass von einem Nischendasein gesprochen werden kann. Dennoch: Die Reichweite der Fake-News-Kanäle selbst ist in Deutschland relativ klein. Sie blieb sowohl im Bundestagswahlkampf 2017, als auch während des Europawahlkampfs 2019 weit hinter der Reichweite von Produkten etablierter Medien zurück (vgl. Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018). Hinzu kommt: In Deutschland greifen soweit, anders als in den USA, nur selten Medien mit großer Reichweite Falschmeldungen von peripheren Fake-News-Seiten auf (Sängerlaub et al., 2018: 3).

In den USA gewonnene Erkenntnisse über Desinformation im Internet können also nicht eins zu eins auf deutschsprachige Fake News übertragen werden. Es ist deshalb notwendig, mit empirischen Studien Erkenntnisse für den deutschsprachigen Raum zu gewinnen, wie dies im vorliegenden Band geschieht. Dabei liegt in der hier beschriebenen Studie der Schwerpunkt der Forschung nicht wie bei anderen Studien darauf, die Reichweite und die Verbreitungswege von Fake News zu untersuchen (Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018), sondern es geht schwerpunktmäßig darum, die Beschaffenheit deutschsprachiger Fake News genau zu analysieren. Entscheidend sind in diesem Zusammenhang folgende Fragen:

Auf welchen Plattformen sind im deutschsprachigen Raum Fake News zu finden? Wie werden hierzulande Fake News produziert und welche journalistischen Selektionsroutinen und Darstellungsformen werden angewandt,

um Aufmerksamkeit zu generieren und Fake News glaubwürdig zu machen? Welche thematischen Schwerpunkte sind auszumachen? Mit welchen Argumentationsmustern versuchen Produzierende von Fake News, ihre Leser zu überzeugen? Und last but not least: Inwieweit können die hier dargestellten Erkenntnisse über die Beschaffenheit von Fake News im deutschsprachigen Raum dazu beitragen, die Verbreitungsmechanismen besser zu verstehen und Fake News effektiver zu bekämpfen?

B. Wer produziert Fake News? Die untersuchten Webseiten

Die empirische Grundlage für die Beantwortung dieser Fragen bietet die Analyse eines Samples von 489 Fake News.[7]

Wie bereits in der Einleitung ausgeführt, werden im vorliegenden Band Fake News als online verbreitete Informationen definiert, die journalistische Nachrichteninhalte nachahmen, indem sie journalistische Routinen der Nachrichtenpräsentation und -auswahl anwenden, bei deren Produktion zugleich aber journalistische Rechercheroutinen systematisch missachtet werden und deshalb falsche Tatsachenbehauptungen enthalten (ähnlich Lazer et al., 2018: 1094).

Viele Definitionen von Fake News betonen, dass der Fehler stets intentional sein muss, der Nachricht also eine Absicht zu desinformieren zugrunde und sich Fake-News-Produzierende zudem bewusst sein müssen, dass sie eine faktisch falsche Information weitergeben (Egelhofer & Lecheler, 2019: 100; Levy, 2017: 20; Shin et al., 2018: 278; für einen Überblick vgl. Zimmermann & Kohring, 2018: 528). Dagegen wenden einige Autoren auf der theoretischen Ebene ein, dass von einer solchen Absicht auszugehen stets eine Zuschreibung ist (Scholl & Völker, 2019: 211). Zudem zeigt sich in empirischen Studien, dass der Nachweis von Vorsätzlichkeit stets eine große Herausforderung darstellt und gerade bei Studien mit großen Datenmengen oftmals nicht erfolgen kann (u.a. Vosoughi et al., 2018) oder behelfsmäßig erfolgen muss, beispielsweise über den Nachweis, dass das Medium keine Richtigstellung veröffentlicht, wenn es auf den jeweiligen Fehler hingewiesen wird (Sängerlaub et al., 2018: 12). Im Rahmen von sozialwissenschaftlichen Studien über Fake News, die mit großen Datenmengen operieren, wie der

7 Das vorgestellte Sample oder einzelne Meldungen aus dem Sample wurde auch für die Untersuchungen anderer Disziplinen verwendet (vgl. insbesondere Kapitel 4, Informatik zur Textstruktur von Fake News, aber auch Kapitel 3, Medienpsychologie).

hier dargestellten Studie, ist es deshalb nicht möglich, Aussagen über die Intention einzelner Fake-News-Produzierenden zu machen.

Dennoch besteht auch in der sozialwissenschaftlichen Beschäftigung mit Fake News die Notwendigkeit, Fake News von „schlechtem Journalismus" (vgl. Wardle, 2017) abzugrenzen, in dem ebenfalls immer wieder Fehler entstehen. Für diese Abgrenzung ist maßgeblich, dass journalistische Fact-Checking-Routinen von Fake-News-Produzierenden systematisch missachtet werden. Dies heißt, Fake News enthalten Informationen, die durch grundlegende und von Journalisten routinemäßig durchgeführte Recherchemethoden wie das Suchen einer zweiten Quelle, das Überprüfen des – angeblichen – Augenzeugen oder einen Anruf bei der ermittelnden Behörde widerlegt werden können. Ob der Fake-News-Produzierende die falsche Tatsachenbehauptung von anderen übernimmt, selbst glaubt und nicht überprüft oder ob er sich voll bewusst ist, dass die von ihm verbreitete Tatsachenbehauptung falsch ist, weil er sie beispielsweise selbst erfunden hat, ist für diese Definition nicht maßgeblich. Maßgeblich ist vielmehr, dass die jeweilige Falschmeldung darauf schließen lässt, dass bei der Produktion zwar journalistische Darstellungsroutinen imitiert wurden, dass aber redaktionelle Normen und Prozesse, die im journalistischen Produktionsprozess für die Korrektheit und Glaubwürdigkeit der Information sorgen, systematisch nicht befolgt wurden (vgl. Lazer et al., 2018: 1094).

Die Erfassung von verdächtigen Inhalten und ihre Überprüfung beschränkte sich auf Beiträge, die von Dezember 2015 bis März 2018 veröffentlicht wurden. Dieser Zeitraum wurde gewählt, weil die Archive der meisten hier untersuchten Webseiten nicht weiter zurückreichen. Darüber hinaus zeigt eine Analyse von Treffern auf Google News, dass in Deutschland bis Dezember 2015 der Begriff „Fake News" kaum verwendet wurde. Innerhalb des genannten Zeitraums finden sich zudem relevante politische Ereignisse wie die Debatte um den Zuzug von Flüchtlingen im Jahr 2015 und die Bundestagswahl 2017 inklusive der daran anschließenden komplizierten Koalitionsverhandlungen.

Die erfassten Meldungen stammen von 39 verschiedenen deutschsprachigen Webseiten (vgl. Tabelle 2.1), die aber teilweise, wie das eingangs erwähnte Portal anonymousnews.ru, außerhalb des deutschsprachigen Raums registriert sind. Aus forschungspraktischen Gründen ist es dabei faktisch unmöglich, ein größeres Sample von Fake News aus einer nach dem Zufallsprinzip erstellten Stichprobe zugewinnen, weil bei einem solchen Vorgehen die Fact-Checking-Prozedur viel zu aufwendig wäre (vgl. Lazer et al., 2018: 1095). Deshalb wurden gezielt Webseiten angesteuert, die in der wissen-

schaftlichen Auseinandersetzung als Fake-News-Seiten aufgeführt werden und auf denen Debunking-Initiativen regelmäßig fündig werden (wie z. B. Schweiger, 2017 oder Mimikama).

Andere Forschungsprojekte zum Thema Fake News im deutschsprachigen Raum haben bereits aufgezeigt, dass es sich bei Desinformation hierzulande um ein primär rechtspopulistisches Phänomen handelt (Humprecht, 2018; Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018). Dennoch wurden bei der Auswahl der beobachteten Seiten zahlreiche Anstrengungen unternommen, über das rechtsgerichtete Milieu und zugleich über „die üblichen Verdächtigen" hinauszugehen: Dafür wurde das Medienumfeld von Organisationen, die im Verfassungsschutzbericht erwähnt werden, systematisch auf Fake News hin untersucht, wobei ein besonderes Augenmerk linken Organisationen galt.

Doch auch bei diesem ergänzenden Zugang wird deutlich, dass Fake News überwiegend im rechtspopulistischen und rechtsextremistischen Milieu zu finden sind.[8] Das vorliegende Sample unterstreicht somit die für kleinere Samples vorliegenden Befunde: Deutschsprachige Fake News wurden im Untersuchungszeitraum zwar nicht ausschließlich, aber doch überwiegend von rechtsgerichteten Medienakteuren produziert. 88,3 Prozent der als Fake News klassifizierten Meldungen wurden auf Webseiten publiziert, die politisch rechts stehen, 8,2 Prozent der Falschmeldungen sind Seiten zuzuordnen, die über keine eindeutige politische Ausrichtung verfügen, sondern Verschwörungstheorien beispielsweise zu Medizinthemen verbreiten und nur 3,5 Prozent der Fake News im Sample stammen von Seiten, die politisch links stehen.[9]

Wichtig ist es anzumerken, dass im deutschsprachigen Raum keine Webseiten existieren, die ausschließlich Falschmeldungen verbreiten: Stattdessen werden auf den untersuchten Seiten auch Beiträge veröffentlicht, in denen die Fakten korrekt wiedergegeben sind, und Meinungsbeiträge, die zwar teilweise extreme Positionen vertreten, aber keine falschen Tatsachenbehauptungen enthalten. Beide Arten von Beiträgen wurden - entsprechend der in Kapitel 1

8 Eine Einschätzung zur Verbreitung durch rechtspopulistische „alternative News-Outlets" bieten auch Holt et al. (2019).

9 Der „Blaue Bote" ist das linke Portal, auf dem die meisten falschen Tatsachenbehauptungen gefunden wurden. Allerdings ist die politische Ausrichtung dieses Portals sehr speziell: Das Portal betont seine politische Neutralität und stellt sich klar gegen „Nazis". Aber es konzentriert sich oftmals auf Außenpolitik-Themen. Hier agiert der „Blaue Bote" stark prorussisch und antiamerikanisch und unterstützt beispielsweise die russische Krim-Annexion: http://blauerbote.com/impressumkontakt/ (Stand: 24.6.2019).

Tabelle 2.1: Liste der Webseiten, von denen die Fake News im Sample stammen

Rechte Portale (%=88,3)	**N**	**Portale mit politisch neutralen Verschwörungstheorien (%=8,2)**	**N**	**Linke Portale (%=3,5)**	**N**
Alles Roger	8	Alpenschau	2	Blauer Bote	13
Anonymous News	15	Philosophia Perennis	12	Blaue Flora/DPR online	2
Bayern Depesche	2	Bereicherungswahrheit	1	Labour Net	2
Berlin Journal	2	Euro-Med	9		
Compact Magazin	5	Signs of the times	16		
Die Unbestechlichen	23				
Epoch Times	20				
Freie Welt	1				
freisleben-news	1				
Gegenfrage	3				
Guido Grandt	27				
Halle Leaks	85				
Info Direkt	9				
Journalisten Watch	35				
MMnews	1				
Netzfrauen	1				
News for Friends	5				
Noch Info	9				
No Islam/Noack Finsterwalde	15				
Opposition24/Freie Presse	20				
Perspektive Online	3				
Politically Incorrect	2				
Rapefugees	15				
RT Deutsch	1				
Schlüsselkind-Blog	15				
Schweizer Morgenpost	44				
Truth24	38				
Unzensuriert	10				
World Socialist Web Site	2				
Zeitenschrift	1				
Zuerst	14				
Gesamt	**432**		**40**		**17**

Anmerkung: Grundlage der Kategorisierung der Portale als rechts-/linkspopulistisches oder Verschwörungstheorie-Portal waren Publikationen von Faktenprüfungsinstitutionen (z. B. Mimikama, http://www.mimikama.at, Stand: 24.6.2019, oder ARD Faktenfinder, http://faktenfinder.tagesschau.de, Stand: 24.6.2019, Schweiger (2017: 48-50) sowie 10000flies https://www.10000flies.de, Stand: 24.6.2019, ein Portal, das u.a. die Reichweite alternativer Social Networks verfolgt. Philosophia Perennis wurde im Untersuchungszeitraum noch als neutral eingestuft, kann aber inzwischen ebenfalls als rechtes Portal eingestuft werden.

dieses Bandes ausgeführten Definition von Fake News – nicht in das Sample aufgenommen. So befinden sich im untersuchten Sample ausschließlich Beiträge, in denen mindestens eine Tatsachenbehauptung identifiziert wurde, die durch journalistische Recherche widerlegt werden konnte.

Bei den aufgenommenen Meldungen wurde systematisch nach weiteren falschen Tatsachenbehauptungen gesucht, wobei bis zu drei falsche Tatsachenbehauptungen einzeln erfasst und zusätzlich eine Einschätzung über den Anteil der kontrafaktischen Inhalte am Gesamttext vorgenommen wurde. Auch hier zeigt sich: Wahrheit und Lüge sind in der deutschsprachigen Fake-News-Szene eng miteinander verwoben[10] (vgl. Tabelle 2.2): Fast die Hälfte der Artikel enthält weniger als 25 Prozent falsche Behauptungen. Nur jeder fünfte Artikel des Samples besteht zu mehr als 50 Prozent aus falschen Behauptungen.

Dies lässt erste Rückschlüsse über die Arbeitsweise der Fake-News-Produzierenden zu: Offenbar werden Fakten, die auch in seriösen Medien zu lesen sind und für die auch oftmals seriöse Quellen angegeben werden, wie die Tagesschau, der Spiegel oder auch Polizeimeldungen, mit Tatsachenbehauptungen vermengt, für die es keinerlei Belege und keine seriösen Quellen gibt (siehe Abb. 2.1). Nähe zu seriöser Berichterstattung wird also simuliert und die falschen Behauptungen sind, bildlich gesprochen, eher giftige Beimischung als Hauptbestandteil von Fake-News-Meldungen.

Zudem wurde ermittelt, in welchem Teil des pseudo-journalistischen Beitrags die falschen Tatsachenbehauptungen jeweils platziert sind, wobei nach Überschrift, Teaser, Fließtext und Foto unterschieden wurde. Interessant ist, dass mehr als die Hälfte der erfassten Meldungen (57,2 %) die falsche Tatsachenbehauptung auch (28,8 %) oder sogar ausschließlich (28,4 %) in Überschrift oder Teaser platzieren – also jenem Teil der Meldung, der darauf abzielt, Leser anzulocken.

Diese Textstruktur legt nahe, dass deutschsprachige Fake-News-Produzierende bereits in den vergangenen Jahren einer Taktik folgten, mit der laut einer aktuellen US-amerikanischen Studie die Reichweite auf Social Networks stark gesteigert werden kann: Vosoughi, Roy und Aral (2018) haben nachgewiesen, dass falsche Tatsachenbehauptungen auf Twitter mit einer um 70 Prozent höheren Wahrscheinlichkeit geteilt werden als wahre, korrekte Neuigkeiten.

10 Vgl. zur Verbindung der Konzepte „Fake News“ und „Lüge“ u.a. auch die Einordnung von Zimmermann und Kohring (2018).

Zahlreiche Feedback-Möglichkeiten ermöglichen es den Betreibern von nachrichtlichen und pseudonachrichtlichen Webseiten und Social-Network-Kanälen, in Echtzeit zu verfolgen, wie erfolgreich sich ein von ihnen veröffentlichter Post oder Artikel verbreitet. Insofern ist es durchaus denkbar, dass das Erfahrungswissen von Fake-News-Produzierenden zu einer Taktik führt, die an die spezifische Gelegenheitsstruktur und Logik von Kommunikationsplattformen angepasst ist (zur „network media logic" vgl. Engesser, Ernst et al., 2017: 1113; Klinger & Svensson, 2015). Unwahre und genau deshalb oft besonders überraschende, skandalöse und konfliktreiche Behauptungen in der Überschrift und im Teaser können dazu genutzt werden, die Aufmerksamkeit von Nutzenden auf einen auf Kommunikationsplattformen verbreiteten Artikel zu ziehen und einen Click-Impuls auszulösen.

Dass in immerhin der Hälfte der so angepriesenen Artikel die falsche Tatsachenbehauptung im eigentlichen Text dann gar nicht mehr erwähnt und in der anderen Hälfte die Tatsachenbehauptung zwar wiederholt, aber meist mit zahlreichen korrekten Fakten vermischt wird, lässt darauf schließen, dass die Texte selbst wiederum eher drauf abzielen, ein gewisses Maß an Seriosität zu vermitteln (vgl. Tabelle 2.2).

C. Fake News als Pseudo-Journalismus – Anwendung journalistischer Selektionsroutinen und Darstellungsformen

In der Literatur wird davon ausgegangen, dass bei der Produktion von Fake News professionelle journalistische Darstellungsroutinen imitiert werden, um bei den Rezipienten Genre-Wissen zu aktivieren und Vertrauen in die Qualität des Produkts zu schaffen (Egelhofer & Lecheler, 2019; Hooffacker & Meier, 2017; Horstmann et al., 2018; Schweiger, 2017). Zugleich grenzen sich viele Fake-News-Produzierende demonstrativ von dem ab, was sie als Mainstream-Medien bezeichnen.[11] Wie stark imitieren Fake News (professionellen/etablierten) Journalismus? Um diese Frage zu beantworten, wird im Folgenden untersucht, welche Darstellungsformen häufig gewählt werden, wie professionell oder unprofessionell die journalistische Anmutung von Fake News im deutschsprachigen Raum ist und welche Selektions- und

11 Inzwischen finden regelmäßig Treffen statt, auf denen auch die Betreiber von vielen der Seiten, auf denen Falschinformationen zu finden sind, vertreten sind. Auf diesen Treffen wird an der Vernetzung und Selbstpositionierung gearbeitet: https://juergenfritz.com/2018/02/25/bloggertreffen-berlin/ (Stand: 24.6.2019).

Tabelle 2.2: Anzahl, Anteil und Positionierung der falschen Tatsachenbehauptungen

Anzahl an Fakes	**Anzahl der Fälle**	**Anteil in %**
1 Fake	171	35,0
2 Fakes	155	31,7
3 Fakes	163	33,3
Gesamt	489	100
Anteil an Fakes		
< 25 %	235	48,1
< 50 %	149	30,5
> 50 %	101	20,7
Unklar	4	0,8
Gesamt	**489**	**100**
Position der Fakes		
Bild	11	2,2 %
Headline/Teaser	139	28,4 %
Fließtext	198	40,5 %
Headline/Teaser und Fließtext	141	28,8 %
Gesamt	**489**	**100**

Basis: N=489 Fake News. Quelle: DORIAN-Sample, eigene Berechnung.

Präsentationskriterien, also welche Nachrichtenfaktoren bei der Produktion von Fake News besonders stark berücksichtigt werden.

Der Blick auf die Darstellungsformen zeigt, dass Fake News in Deutschland sehr weitgehend auf nachrichtliche Darstellungsformen beschränkt sind (vgl. Tabelle 2.3): Bei fast 34 Prozent handelt es sich um kurze Nachrichten, bei 48 Prozent um längere Berichte. Die Darstellungsform des angefeatureden Berichts, die im seriösen Journalismus inzwischen häufig verwendet wird (Flath, 2013: 62ff.), wird im Rahmen der Fake-News-Produktion nur sehr selten imitiert. Dies zeigt, dass Fake News darauf ausgerichtet sind, als „Nachrichten" im engeren Sinn verstanden zu werden. Stärker erzählerische journalistische Darstellungsformen werden deshalb nicht angewandt.[12]

Um Aussagen über das Maß der Professionalität der Meldungen im Sample zu treffen, wurde zunächst untersucht, inwieweit die Texte Schwächen

12 Um Emotionen auszulösen, wären reportage-ähnliche Erzählformen an sich geeignet. Denkbar ist, dass diese nicht gewählt werden, weil sie aufwendig zu produzieren sind und viele journalistische Kenntnisse voraussetzen.

Tabelle 2.3: Anteile der verschiedenen Darstellungsformen

Darstellungsform	Anzahl Fälle	Anteil in %
Bericht	236	48,3
Nachricht	166	33,9
Mischformen	63	12,9
Angefeatureder Bericht	24	4,9
Gesamt	**489**	**100**

Basis: N=489 Fake News. Mischformen stellen uneindeutige Formate dar, etwa Berichte, die vom Aufbau her Nachrichten ähneln, aber zu lang für die Einordnung als klassische Nachricht sind. Quelle: DORIAN-Sample, eigene Berechnung.

enthalten, die in professionellen journalistischen Texten soweit als möglich vermieden werden (vgl. Tabelle 2.4). Tatsächlich enthalten 21 Prozent der erfassten Fake News mehr als zwei Rechtschreibfehler, 20 Prozent eine große Zahl von Fehlern bei der Interpunktion und 13 Prozent der Meldungen eine für etablierten Journalismus untypisch hohe Verwendung von Majuskeln.[13] Viele der Texte weisen also offensichtliche professionelle Mängel auf. Zugleich zeigen diese Zahlen aber auch, dass rein formale Fehler kein geeigneter Indikator sind, um Fake News im deutschsprachigen Raum zu erkennen, weil sie in den meisten Beiträgen vermieden werden können.

Was den Umgang mit Fakten betrifft, sind die Unterschiede jedoch schon deutlicher: Eine journalistische Grundregel besagt, dass Vermutungen, die durch keinerlei Daten zu belegen sind und für die es auch keine Quelle gibt, vermieden werden sollten. 55 Prozent der untersuchten Fake News weisen allerdings solche Vermutungen komplett ohne jeden Beleg auf.[14]

13 Majuskeln, also die durchgängige Großschreibung einzelner Wörter, ist in der nichtjournalistischen Onlinekommunikation zur Hervorhebung oder um Wut auszudrücken durchaus üblichen, von Journalistinnen und Journalisten wird sie jedoch weitgehend vermieden. Fehler in Rechtschreibung und Zeichensetzung sind wiederum auch in journalistischen Texten anzutreffen – verstärkt seitdem die Schlusskorrektur in vielen Redaktionen eingespart wurde.

14 Ein Beispiel für eine solche Vermutung (und in diesem Fall auch Anschuldigung), die auf keinerlei Quelle zurückgeführt wird, wäre folgende Formulierung: „Wären die Zahlen hinter den zahlreichen Statistikmanipulationen des BKAs bekannt, würde die Straftaten durch Flüchtlinge gewiss längst die Millionengrenze überschritten haben.“ https://www.journalistenwatch.com/2017/09/22/bka-studie-fluechtlinge-begehen-628-000-straftaten/ (Meldung vom 22.9.2017, Stand: 24.6.2019)

Tabelle 2.4: Journalistische Professionalitätsindikatoren I

Indikator journalistischer Professionalität	Anzahl Fälle	Anteil in %
Mutmaßungen/Vermutungen	270	55,2
Orthografiedefizite	104	21,3
Gehäufte Interpunktionsfehler	99	20,2
Verwendung von Majuskeln	65	13,3
Anonymisierte Quellennennung	43	8,8

Basis: N=489 Fake News. Quelle: DORIAN-Sample, eigene Berechnung.

Auf anonyme Quellen wird hingegen nur recht selten, in neun Prozent der Meldungen, verwiesen.[15]

Darüber hinaus wurden komplexere Professionalitätsindikatoren wie die Faktenkonsistenz innerhalb einer Meldung, sprachliche Präzision, ein stringenter Textaufbau, die Qualität des Leadsatzes oder auch Professionalität der Überschrift auf einer fünfstufigen Skala erfasst (vgl. Tabelle 2.5). Den Maßstab für professionelles Handeln stellen dabei Regeln dar, die in gängigen Lehrbüchern für praktischen Journalismus in Deutschland vermittelt werden, wobei bevorzugt auf Texte über professionellen Onlinejournalismus zugegriffen wurde (vgl. LaRoche et al., 2013; Liesem, 2015; Plöchinger, 2014).

Hier zeigt sich, dass der Professionalitätsgrad von Fake-News-Seiten im deutschsprachigen Raum niedrig ist. Die Seiten imitieren zwar das äußere Erscheinungsbild von journalistischen Texten und vermeiden grobe Fehler, die Missachtung grundlegender Darstellungsregeln für professionellen Journalismus ist jedoch eher die Regel als die Ausnahme: In einem großen Teil der Meldungen sind die Fakten nicht konsistent, ist die Sprache nicht präzise und werden die Regeln zum journalistischen Textaufbau nicht berücksichtigt, die besagen, zuerst das Aktuelle und Relevante darzustellen und dann erst auf Vorgeschichte und Hintergründe einzugehen (vgl. zur journalistischen Text-

15 Im Journalismus ist es üblich bei Vermutungen, deren Quelle nicht namentlich zitiert werden kann, zumindest einen Hinweis auf die Art der Quelle anzufügen, wie beispielsweise „aus Regierungskreisen“. Weil solche Aussagen nur schwer überprüfbar sind, war zu vermuten, dass Fake-News-Produzierende stark auf solche Redewendungen zurückgreifen. Tatsächlich findet diese Art der Quellen-Verschleierung im Bereich der deutschsprachigen Fake News soweit nur in Einzelfällen statt.

Tabelle 2.5: Journalistische Professionalitätsindikatoren II (skaliert)

Journalistische Professionalitätsindikatoren	Anzahl Fälle, die als sehr professionell oder professionell eingestuft wurden	Anteil in %
Überschrift	185	38 %
Leadsatz	166	34 %
Faktenkonsistenz	139	29 %
Textaufbau, der journalistische Relevanz-Regeln berücksichtigt	132	27 %
Sprachliche Präzision	116	24 %

Basis: Vollständiges Sample. N = 489. Fünfstufige Skala (0=sehr unprofessionell, 4=sehr professionell), bei der jeweils die beiden stärksten Werte hier aufgeführt werden.

struktur z. B. Flath, 2013). Eine Textstruktur, die bei wenig professionellen Fake News häufig zu finden ist, weist Abb. 2.4 auf Seite 62.

Auffällig ist, dass Überschrift und Leadsatz, also die Teile der Meldungen, die zur Aufmerksamkeitsgenerierung beitragen, noch am häufigsten als sehr professionell oder professionell zu bewerten sind, nämlich in einem guten Drittel der Fälle. Daraus kann geschlossen werden, dass Fake-News-Macher sich hier noch am ehesten um ein professionelles Erscheinungsbild bemühen, die eigentlichen Fließtexte aber ohne Sorgfalt zusammengeschrieben oder sogar zusammenkopiert werden.[16]

Komplexere Professionalitätsindikatoren können derzeit also noch dazu beitragen, Fake News als solche zu erkennen. Allerdings ist damit zu rechnen, dass sich zumindest größere Fake- News-Seiten in diesen Dimensionen professionalisieren werden.

Doch wie hängt der Professionalitätsgrad der Meldungen mit der Dichte der falschen Tatsachenbehauptungen zusammen? Hierfür wurden die bereits aufgeführten Professionalitätsindikatoren neu skaliert (sehr unprofessionell=geringe Professionalität, unprofessionell bis professionell=Schwächen, sehr professionell=hohe Professionalität).

16 Studien zeigen, dass viele Social-Networks-Nutzende Beiträge teilen, von denen sie nur die Überschrift kennen (Gabielkov et al., 2016). Insofern liegt der Schluss nahe, dass dieser Prioritätensetzung der Fake-News-Produzierenden einer Kosten-Nutzen-Maximierung in der „network media logic“ zugrunde liegt, vgl. Abschnitt F.

Tabelle 2.6: Zusammenhang zwischen Fake-Anzahl, Fake-Anteil und Professionalitätsindikatoren

	Professionalität der Fake News (Anteil in %)			
Anzahl an Fakes	**Hoch (16,4 %)**	**Schwächen (14,1 %)**	**Gering (69,4 %)**	**Gesamt (in %)**
1 Fake	22,8	19,2	58,1	100
2 Fakes	18,5	14,6	66,9	100
3 Fakes	8,0	8,6	83,4	100
Gesamt	**79**	**68**	**334**	**100**
Anteil Fakes				
< 25 %	22,0	19,8	58,2	100
< 50 %	11,5	9,5	79,1	100
> 50 %	11,2	6,1	82,7	100
Gesamt	**79**	**66**	**333**	**100**

Basis: N=481 Fake News (Anzahl an Fakes) bzw. N= 478 Fake News (Anteil an Fakes). Fehlende Fälle stellen nicht eindeutig zuordenbare Fälle auf den drei Teildimensionen dar. Anzahl an falschen Tatsachenbehauptungen: X^2=25,964, df=4, p<0,001; Cramers V=0,164, p<0,001. Anteil Fakes: X^2=29,060, df=4, p<0,001; Cramers V=0,174, p<0,001.

Die in Tabelle 2.6 dargestellte Auswertung zeigt, dass Meldungen mit einem geringen Professionalitätsgrad oftmals viele falsche Behauptungen enthalten, während Meldungen mit einem hohen Professionalitätsgrad oftmals nur eine falsche Tatsachenbehauptung enthalten. Dieser Zusammenhang kann auf zwei Arten interpretiert werden: Möglich ist, dass Artikel, die weniger Lügen enthalten, von professionelleren Produzierenden erstellt werden, die dem Journalismus näherstehen und journalistische Regeln besser kennen. Möglich ist aber auch, dass der Zusammenhang schlicht darauf zurückzuführen ist, dass viele Falschmeldungen aus Textbausteinen erstellt werden, die aus Polizeimeldungen, Pressemitteilungen und seriösen Medienberichten stammen und mit falschen Tatsachenbehauptungen angereichert werden. Es ist also denkbar, dass beim Einbauen von falschen Tatsachenbehauptungen jeweils ein Stück professioneller Textqualität verloren geht.

Eine zentrale Rolle in der journalistischen Nachrichtenproduktion spielen Nachrichtenfaktoren: Sie helfen Journalisten und Journalistinnen einerseits zu entscheiden, ob ein Ereignis berichtenswert ist, andererseits helfen sie, Meldungen zu strukturieren, weil Teilaspekte, die als besonders relevant

betrachtet werden, bei der Darstellung des Ereignisses prominent platziert werden. Die Berücksichtigung von Nachrichtenfaktoren wie beispielsweise Reichweite, negative Folgen oder Nähe stellt also eine zentrale Routine im journalistischen Produktionsprozess dar, was in der Journalistik seit mehr als 70 Jahren immer wieder untersucht wird (Galtung & Ruge, 1965; Kepplinger, 2008; Warren, 1934). Neuere Studien beschäftigen sich in diesem Zusammenhang auch damit, welche Nachrichtenfaktoren in nutzergenerierten Auswahlen und in Kommunikationsplattformen eine besondere große Rolle spielen (z. B. Lee, 2009; Wendelin et al., 2014).

Um herauszufinden, in welchem Maße Fake News Journalismus imitieren und ob Journalismus und Fake News anhand struktureller Merkmale unterscheidbar sind, ist es logisch, auch zu untersuchen, in welchem Ausmaß Fake News Nachrichtenfaktoren berücksichtigen und welche Nachrichtenfaktoren bei der Produktion von Fake News eine besonders wichtige Rolle spielen. Tabelle 2.7 zeigt, dass Nachrichtenfaktoren bei der Produktion von Fake News stark berücksichtigt werden. Einige Nachrichtenfaktoren, wie beispielsweise Nähe, Negativität und Etablierung, spielen bei Fake News eine besonders große Rolle und finden bei jeweils über 80 Prozent der Meldungen Berücksichtigung.[17]

Dieser Schwerpunkt entspricht allerdings weder den zentralen Nachrichtenfaktoren, die für klassischen Journalismus ausfindig gemacht wurden (Ruhrmann & Göbbel, 2007), noch dem, was neueren Studien zufolge in nutzergenerierten Nachrichten-Auswahlen oder auf Social Networks eine zentrale Rolle spielt (vgl. Bednarek & Caple, 2017; Wendelin et al., 2014). Die starke Konzentration auf die Nachrichtenfaktoren Nähe, Negativität und Etablierung scheint also spezifisch für die hier untersuchten Fake News zu sein, was jedoch mit ihrer thematischen Ausrichtung zusammenhängen könnte, die im nächsten Teilkapitel untersucht wird.

17 Ob Fake-News-Produzierende sich bewusst oder unbewusst an Nachrichtenfaktoren ausrichten, kann im Rahmen dieser Studie nicht nachgewiesen werden. Weil Nachrichtenfaktoren an grundlegenden menschlichen Wahrnehmungsmustern orientiert sind, könnte die starke Berücksichtigung der Nachrichtenfaktoren auch schlicht auf Erfahrungswerte und die kontinuierliche Auswertung von Klickzahlen zurückgeführt werden.

Tabelle 2.7: Ausmaß/Anteil der verwendeten Nachrichtenfaktoren

Nachrichtenfaktor	Anzahl Fälle	Anteil in %
Nähe	431	88,1
Negativität	413	84,5
Etablierte Themen	402	82,2
Schaden	343	70,1
Kontroverse	322	65,8
Prominenz	216	44,2
Emotion	166	33,9

Basis: N=489 Fake News. Quelle: DORIAN-Sample, eigene Berechnung.

D. Migration und innere Sicherheit - Thematische Schwerpunkte deutschsprachiger Fake News

Bei Betrachtung der thematischen Schwerpunkte von Fake News im deutschsprachigen Raum wird deutlich, dass die Themen innere Sicherheit und Migration im Untersuchungszeitraum die Agenda deutschsprachiger Fake News klar dominieren (vgl. Tabelle 2.8).

Dies wird noch deutlicher, wenn man jene Meldungen betrachtet, die kombiniert, also mittels Haupt- und Nebenthema, in einer Meldung über diese beiden Aspekte berichten (35,4 %). Ein weiteres Drittel der Meldungen berichtet ausschließlich über eines der beiden Themen Migration (15,5 %) und innere Sicherheit (12,9 %) (vgl. Tabelle 2.9). Alle weiteren Themen werden weit weniger intensiv behandelt. Dies entspricht dem, was Humprecht (2018) in ihrer vergleichenden Studie über Fake News im deutsch- und englischsprachigen Raum herausgefunden hat: Während englischsprachige Fake News sich primär an politischen Akteuren abarbeiten, zielen deutschsprachige Fake News vor allem auf Migranten ab und bringen diese mit kriminellen Aktivitäten in Verbindung (Humprecht, 2018: 9). Auch hier könnte eine Anpassung der Fake-News-Produzierenden an die spezifische Verbreitungslogik im digitalen Umfeld vorliegen, zumal Kriminalitätsthemen von deutschsprachigen Internet-Nutzenden allgemein stark nachgefragt und weiterempfohlen werden (vgl. Wendelin et al., 2014: 452).

Vor dem Hintergrund dieser thematischen Ausrichtung ist die Konzentration auf den Nachrichtenfaktor Nähe besonders interessant, zumal dieser Nachrichtenfaktor in Falschmeldungen zum Thema Migration und Sicherheit signifikant häufiger zu finden ist als in Meldungen mit anderen thematischen

Tabelle 2.8: Themenschwerpunkte von Fake News im deutschsprachigen Raum

Thema	Anzahl Fälle	Anteil in % Hauptthema	Anzahl Fälle	Anteil in % Nebenthema
Innere Sicherheit	208	42,5	50	10,2
Migration	85	17,4	217	44,4
Justiz	42	8,6	45	9,2
Medien	32	6,5	5	1,0
Internationale Beziehungen	26	5,3	23	4,7
Sozialpolitik	21	4,3	17	3,5
Sonstige Themen	75	15,3	40	8,2
Gesamt	**489**	**100**	**489**	**100**

Basis: N=489 Fake News. Sonstige Themen beinhaltet Arbeitsmarkt-, Wirtschafts-, Bildungs-, Finanz-, Kultur- und Europapolitik. Quelle: DORIAN-Sample, eigene Berechnung.

Schwerpunkten (vgl. Tabelle 2.10). Auch der Nachrichtenfaktor Schaden ist in Meldungen zum Thema Migration und Sicherheit signifikant öfter vertreten.

Hier wird ein grundlegendes Muster deutschsprachiger Fake News deutlich: Ein großer Teil der Fake News zielt darauf ab, den Eindruck zu verbreiten, dass „Flüchtlinge“ und „Messermigranten“[18] Gefahr und Schaden in die direkte Umwelt der Rezipientinnen und Rezipienten tragen. Dabei wird aufgezeigt, dass die Opfer dieser angeblichen und mitunter auch realen, aber dennoch mit falschen Tatsachenbehauptungen ausgeschmückten Gewaltattacken den Rezipientinnen und Rezipienten geographisch und auch kulturell nahestehen, was impliziert, dass der Schaden auch sie selbst treffen könnte. Die eingangs am Beispiel von anonymousnews.ru beschriebene Taktik, durch die Schilderung von Verbrechen Angst zu verbreiten und Migranten auszugrenzen, ist also nicht nur bei dieser Seite zu finden, sondern stellt ein im gesamten Sample häufig anzutreffendes Muster dar.

Tabelle 2.10 zeigt darüber hinaus, dass die Themen Migration und innere Sicherheit – anders als alle anderen Themen – nur sehr selten als Kontroverse

18 Ein Beispiel für den Versuch, die Bezeichnung „Messermigranten“ medial zu etablieren, findet sich unter https://www.lr-online.de/nachrichten/brandenburg/messermigranten-bedrohen-das-land-laut-afd-fraktionschefin-alice-weidel-hat-sich-die-sicherheitslage-in-deutschland-dramatisch-verschaerft-_aid-33067619 (Stand: 24.6.2019).

Tabelle 2.9: Kombinierte Themenschwerpunkte von Fake News

Thema	Anzahl Fälle	Anteil in %
Migration & Innere Sicherheit	173	35,4
Migration	76	15,5
Innere Sicherheit	63	12,9
Justiz	42	8,6
Medien	32	6,5
Internationale Beziehungen	26	5,3
Sozialpolitik	21	4,3
Sonstige Themen	56	11,5
Gesamt	**489**	**100**

Basis: N=489 Fake News. Sonstige Themen beinhaltet Arbeitsmarkt-, Wirtschafts-, Bildungs, Finanz-, Kultur- und Europapolitik. Quelle: DORIAN-Sample, eigene Berechnung.

dargestellt werden. Auch dieser Unterschied ist signifikant. Während Fake-News-Produzierende bei anderen Themen durchaus die Kontroverse zwischen unterschiedlichen Akteuren abbilden, um den Artikel interessant zu machen, wird bei diesen Themen nur sehr selten eine Diskussion mit unterschiedlichen Standpunkten –zum Beispiel über die genauen Umstände und Hintergründe der Tat – abgebildet.

Humprecht (2018), die für den deutschsprachigen Raum ebenfalls den Themenschwerpunkt Immigration aufzeigt, weist in ihrer vergleichenden Studie darauf hin, dass die thematische Agenda von Fake News eng mit der thematischen Agenda der seriösen Medien im jeweiligen Land verbunden ist. Sie arbeitet heraus, dass Gesundheitsthemen in den USA während der Debatte um verpflichtende und bezahlbare Krankenversicherung auch die Agenda von Fake News dominierten, dort gefolgt vom Thema „Kriminalität und Justiz". Das Thema Migration spielt in den USA in Humprechts Untersuchungszeitraum hingegen eine geringere Rolle (2018: 9). Sie berücksichtigt zwar nicht den Einsatz von Nachrichtenfaktoren, die von ihr als dominant herausgearbeiteten Themen für die US-Fake-News eignen sich jedoch ebenfalls besonders gut, um durch eine Betonung von Nachrichtenfaktoren wie Nähe und Schaden ein Angstnarrativ aufzubauen.

Bei Betrachtung des Anteils der falschen Tatsachenbehauptungen speziell auf diesen thematischen Schwerpunkt bezogen fällt auf, dass Fake News zum Thema Migration und Sicherheit tendenziell mit einem geringeren Anteil

Tabelle 2.10: Thematisch ausgerichtete Verwendung von Nachrichtenfaktoren

	Themenschwerpunkte			
Nachrichtenfaktor	**Migration & Sicherheit, Anteil in %**	**Andere Themen, Anteil in %**	**Differenz**	
Schaden	75,6	60,5	+15,1	***
Nähe	92,3	80,8	+11,5	***
Negativität	87,5	79,1	+8,4	*
Emotion	35,3	31,6	+3,7	
Etablierte Themen	80,1	85,9	-5,8	
Prominenz	39,4	52,5	-13,1	**
Kontroverse	58,7	78,8	-20,1	***

Basis: N=489 Fake News. Signifikanztests beruhen auf Chi-Quadrattests
***=p<0,001, **=p<0,01, *=p<0,05. Quelle: DORIAN-Sample, eigene Berechnung.

von falschen Tatsachenbehauptungen operieren als Fake News zu anderen Themen (vgl. Tabelle 2.11).

Gerade im Themenfeld Migration und innere Sicherheit basiert also die Desinformationsstrategie nicht primär darauf, Vorfälle komplett zu erfinden, sondern es wird durch die Zusammenstellung der Vorfälle und das Hinzufügen erfundener Details ein bedrohlicher Eindruck erweckt (vgl. Abb. 2.2).

E. Populistische Argumentationsmuster in deutschsprachiger Desinformation

Eine weitere Möglichkeit, charakteristische argumentative Strategien von Fake News herauszuarbeiten, bietet ihre Untersuchung auf populistische Muster hin.

Populismus wird als eine Ideologie verstanden, die einen Gegensatz zwischen einem reinen und in sich homogenen Volk und einer korrupten Elite konstruiert (vgl. Mudde, 2004; 2016). Populistische Politiker nehmen für sich in Anspruch, als einzige auf der Seite des Volkes zu stehen und leiten aus dieser Positionierung den Anspruch ab, den Willen des Volkes und damit die Wahrheit zu erkennen. Gesellschaftliche Aushandlungsprozesse zwischen verschiedenen, jeweils autonom agierenden Institutionen wie Parteien, Ge-

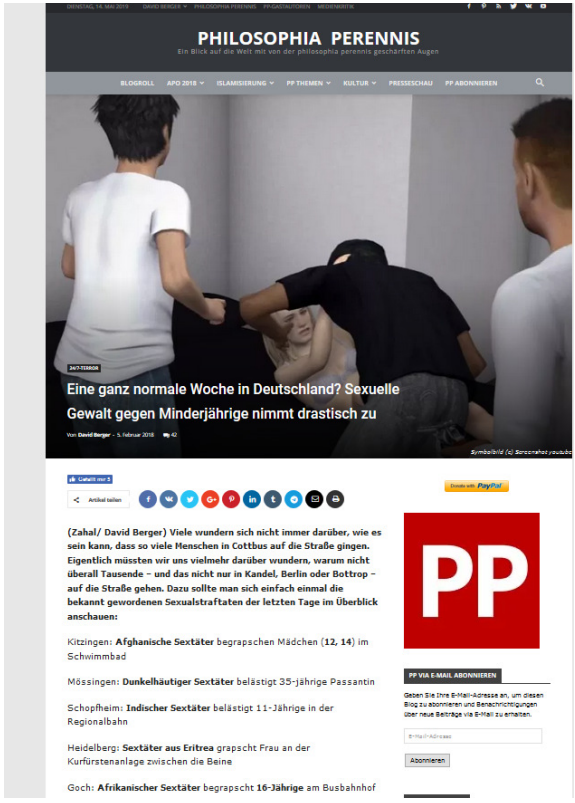

Abbildung 2.2: Migration und Kriminalität – Mischung von korrekten Fakten und falschen Tatsachenbehauptungen

Manche der aufgeführten Schlagzeilen geben Polizeimeldungen korrekt wieder, bei anderen wurden Details hinzuerfunden – beispielsweise fand in einem im Artikel aufgeführten Fall in Wiesbaden zwar ein Überfall, aber kein sexueller Übergriff statt. Die in der Überschrift geäußerte Tatsachenbehauptung, dass die sexuelle Gewalt gegen Minderjährige drastisch zunimmt, wird durch keine Kriminalstatistik belegt. 2017 nahmen die Übergriffe sogar leicht ab. Quelle: https://philosophia-perennis.com/2018/02/05/normale-woche/ (Stand: 24.6.2019).
Quelle Debunking: https://beauftragter-missbrauch.de/presse-service/pressemitteilungen/detail/zahlen-der-polizeilichen-kriminalstatistik-zeichnen-ein-trauriges-bild (Stand: 24.6.2019).

Tabelle 2.11: Zusammenhang zwischen Fake-Anzahl, Fake-Anteil und thematischer Ausrichtung

	Themenschwerpunkte			
	Migration & Sicherheit		Andere Themen	
Anzahl an Fakes	**Anzahl Fälle**	**Anteil in %**	**Anzahl Fälle**	**Anteil in %**
1 Fake	114	66,7	57	33,3
2 Fakes	94	60,6	61	39,4
3 Fakes	104	63,8	59	36,2
Gesamt	**312**	**63,8**	**177**	**36,2**
Anteil Fakes				
25 %	164	69,8	71	30,2
50 %	93	62,4	56	37,6
50 %	54	53,5	47	46,5
Gesamt	**312**	**63,8**	**177**	**100**

Basis: N=489 Fake News. Unklare Fälle zum Fake-Anteil nicht ausgewiesen (N=4). Anzahl an Fakes: X^2=1,276, df=2, p=0,528; Cramers V=0,051, p=0,528. Anteil Fakes: X^2=11,050, df=3, p<0,05; Cramers V=0,150, p<0,05.

werkschaften, Medien und Universitäten halten sie hingegen für nicht der Wahrheitsfindung dienlich (vgl. Tumber & Waisbord, 2017; Waisbord, 2018).

Eine verwandte Haltung kann bei den Betreibern von Webseiten, die im deutschsprachigen Raum Fake News verbreiten, festgestellt werden: Viele von ihnen kooperieren ganz offen mit populistischen Parteien wie der Alternative für Deutschland (AfD). Sie erklären sich dabei mit der AfD nicht nur in inhaltlichen Fragen solidarisch, sondern veranstalten auch gemeinsame Treffen im Bundestag. Auf Grundlage eines nicht-populistischen Politikverständnisses müsste man dies als parteiisches, abhängiges Verhalten betrachten. Die Fake-News-Produzierenden leiten aber genau aus dieser Nähe zur AfD ab, dass sie „freie Medien" seien.[19] Frei ist demnach also nicht, wer sich überparteilich positioniert, sondern wer dem populistischen Weltbild gemäß

19 Zur Selbstpositionierung deutscher Fake-News-Produzierender und ihrer Kooperation mit der AfD: https://philosophia-perennis.com/2019/05/12/der-erste-kongress-der-freien-medien-im-bundestag-eine-sternstunde-des-journalismus/ (Stand: 24.6.2019). Berichterstattung der Zeit über den Kongress: https://www.zeit.de/politik/deutschland/2019-05/afd-bundestag-konferenz-freie-medien-blogger (Stand: 24.6.2019)

auf der Seite des Volks und auf Seite der populistischen Politiker steht. In den USA werden so positionierte Medienakteure als „Hyper-Partisan-Media" bezeichnet und dort inzwischen als einflussreicher Faktor im politischen Geschäft betrachtet (vgl. Benkler et al., 2018: 69-70).

In Deutschland stellen die Betreiberinnen von Seiten, auf denen nachweislich Fake News veröffentlicht werden, an die AfD selbstbewusst Forderungen: So verlangen sie mehr Shares ihrer Artikel auf Social-Network-Accounts von AfD-Parteigrößen, weil das für sie mehr Einfluss und zudem eine geldwerte Erhöhung des Social Involvements bedeute. Ihren Anspruch auf bevorzugte Verbreitung leiten sie dabei daraus ab, dass sie dem „Gemeinwohl" dienlich seien.[20]

Fake-News-Produzierende im deutschsprachigen Raum sind also keine klassischen Medienakteure, sondern vielmehr politisch-mediale Akteure, die eine eigene politische und ökonomische Agenda verfolgen und zugleich populistischen Politikerinnen und Politikern nahestehen. In ihrer Selbstbeschreibung bringen sie zudem ein populistisches Weltbild zum Ausdruck. Es macht daher Sinn, zu untersuchen, inwieweit dieses Weltbild auch in den von ihnen produzierten Fake News Ausdruck findet.

Hierzu eignen sich aktuelle Ansätze der Populismus-Forschung, die darauf abzielen, Populismus nicht ausschließlich als eine Ideologie politischer Akteure zu begreifen, sondern darüber hinaus als kommunikative Strategie zu verstehen, die in konkreten Aussagen messbar wird – egal ob diese Aussagen nun von Politikern, medialen Akteuren oder politisch-medialen Akteuren getätigt wurden (Aalberg et al., 2017; Blassnig et al., 2018; Engesser, Fawzi et al., 2017; Reinemann et al., 2019; Vreese et al., 2018). Mit Hilfe dieser Ansätze lässt sich zudem herausarbeiten, welche Typen populistischer Kommunikationsstrategien in deutschsprachigen Fake News besonders häufig anzutreffen sind. Dabei wird eine Populismus-Typologie von Jagers und Walgrave (2007) angewandt.

Für die vorliegende Studie wurden drei Dimensionen von Populismus erfasst: Eine argumentative Bezugnahme auf das (einfache) Volk, wobei das Volk als homogene Gruppe dargestellt wird, die per se gute, gemeinsame Ziele verfolgt (1), gegen Eliten gerichtete Kommunikationsstrategien, wobei die Eliten als korrupt und einheitlich dargestellt werden (2), gegen ethische,

20 Zum Selbstverständnis deutschsprachiger Fake-News-Produzierender und ihrem Anspruch, von AfD-Politikern bevorzugt verbreitet zu werden, da dies dem „Gemeinwohl" dienlich sei: https://juergenfritz.com/2019/05/12/afd-puscht-mainstreammedien/ (Stand: 24.6.19).

religiöse, kulturelle oder auch sexuelle Minderheiten gerichtete Kommunikationsstrategien, in denen Minderheitsgruppen als nicht zum homogenen Volk gehörend diskriminiert werden (3) (vgl. Aalberg et al., 2017; Jagers & Walgrave, 2007; Reinemann et al., 2019; Vreese et al., 2018).

Diesem Populismus-Verständnis zufolge muss zwingend die erste Bedingung erfüllt sein, damit von Populismus gesprochen werden kann: Ohne eine explizite Bezugnahme auf das Volk, dem positive Eigenschaften und homogene Interessen oder auch gemeinsame Nöte zugeschrieben werden, liegt kein Populismus vor (vgl. Jagers & Walgrave, 2007: 323, ähnlich auch: Mudde, 2004; Mudde & Rovira Kaltwasser, 2016). Zugleich ist diese grundlegende Form des Populismus in der politischen Kommunikation – auch in der politischen Kommunikation von nicht-populistischen Parteien – extrem verbreitet. Studien belegen, dass gerade in Kommunikationskanälen wie Social Networks, die zur Kürze und Vereinfachung zwingen und zugleich eine direkte Kommunikation zwischen Politikerinnen und Politikern mit Wählenden möglich machen, politische Akteure aller Parteien regelmäßig populistische Argumentationsmuster dieser Prägung bemühen (Engesser, Fawzi et al., 2017; Jagers & Walgrave, 2007). Jagers und Walgrave bezeichnen diese milde Form des Populismus als „empty populism", also leeren Populismus. Werden zudem noch Minderheiten ausgrenzt ist von „excluding populism", also ausgrenzendem Populismus, die Rede. Wenn sich die populistische Kommunikation nicht gegen Minderheiten, aber gegen Eliten richtet, sprechen Jagers und Walgrave von „anti-elitist populism" also Anti-Eliten-Populismus. Wenn Minderheiten und Eliten zugleich stigmatisiert werden, ist von „complete populism", also vollständigem Populismus die Rede (Jagers & Walgrave, 2007: 334).

Die Untersuchung des Fake-News-Samples bestätigt eine starke Verbindung zwischen Fake News und Populismus: 75,1 Prozent der analysierten Fake News wenden populistische Kommunikationsstrategien an, wobei nur 12 Prozent der Meldungen den in der politischen Online-Kommunikation auch bei demokratischen Kräften sehr weit verbreiteten leeren Populismus anwenden (thin populism) – d.h. in 63 Prozent der Fake News werden starke populistische Argumentationsmuster sichtbar (thick populism), die auf Ausgrenzung abzielen (Tabelle 2.11). Der Zusammenhang zwischen Populismus und Fake News, der so oft vermutet wird, kann hier also nachgewiesen werden: Die untersuchten Fake News im deutschsprachigen Raum sind großteils von populistischen Kommunikationsstrategien geprägt.

Zudem ist auffällig, dass trotz des thematischen Schwerpunkts auf Migrations- und Sicherheitsthemen Populismus mit anti-elitären Zügen

Tabelle 2.12: Populismustypen in Fake News

Populismus vorhanden	Populismustypus	Anzahl Fälle	Anteil in %
Ja	Leerer Populismus	59	12,1
	Minderheiten ausgrenzender Populismus	93	19,0
	Anti-Eliten-Populismus	128	26,2
	vollständiger Populismus	87	17,8
Nein	Ausgrenzende Kommunikation ohne Volksbezug	38	7,8
	Keine populistische Aussage enthalten	84	17,2
Gesamt		**489**	**100**

Basis: N=489 Fake News. Populismustypen angelehnt an Jagers und Walgrave (2007) und kategorisiert mittels hierarchischer Clusteranalyse. Quelle: DORIAN-Sample, eigene Berechnung.

noch stärker angewendet wird als jener Populismus mit gegen Minderheiten gerichteten Argumentationsmustern: 26,2 Prozent der Meldungen sind dem Anti-Eliten-Populismus zuzuordnen, 19 Prozent dem Minderheiten-ausgrenzenden-Populismus und 17,8 Prozent dem vollständigen Populismus, der beide Gruppen stigmatisiert (vgl. Tabelle 2.12).

Dies zeigt, dass Fake News im deutschsprachigen Raum einen starken politischen Impetus haben: In deutschsprachigen Fake News geht es in hohem Maße darum, bestehende Eliten zu diskreditieren, ihnen die Schuld für verschiedene Probleme zuzuschreiben und das Vertrauen in Elitenakteure unterschiedlicher Bereiche zu untergraben. Die Angst vor „Flüchtlingen" und von diesen ausgehender Kriminalität stellt dabei oftmals ein Mittel zum Zweck dar (vgl. Abb. 2.3)

Ausgrenzende Kommunikationsstrategien finden sich darüber hinaus auch noch in einer kleinen Anzahl von Fake News, die keinen direkten Bezug zum Volk bzw. zu einem einheitlichen Willen oder Interesse des Volkes herstellen. So wird in 7,8 Prozent der Falschmeldungen im Sample ohne Volksbezug gegen Eliten und/oder Minderheiten argumentiert. Insgesamt beinhalten also über 70 Prozent der Fake News explizit ausgrenzende Argumentationsstrategien. Dies beweist, dass Fake News nicht nur polarisieren, indem sie in Frage stellen, was als gesicherte Tatsache gelten kann, sondern

Abbildung 2.3: Anti-Elitenpopulismus in Fake News zu Migration und Kriminalität

Das Beispiel zeigt, wie Angst vor Migranten und von Migranten ausgehender Kriminalität genutzt wird, um Eliten zu diskreditieren. Hier wird beispielsweise das Gatestone-Institut folgendermaßen zitiert:
„Das Anschwellen der Stichwaffengewalt in Deutschland fällt zusammen mit der Entscheidung von Bundeskanzlerin Angela Merkel, rund zwei Millionen Migranten aus Afrika, Asien und dem Nahen Osten ins Land zu lassen. Die Zahl der angezeigten Messerstraftaten ist in Deutschland in den letzten vier Jahren um 600 Prozent in erschreckendem Ausmaß in die Höhe geschnellt – von rund 550 im Jahr 2013 auf fast 4.000 im Jahr 2016."
Tatsächlich ist das Gatestone-Institut als Quelle für Falschmeldungen international bekannt. Die hier zitierten Zahlen beruhen auf keiner Kriminal-Statistik. Mit dem Messer begangene Gewalttaten werden in Kriminalstatistiken nicht getrennt erfasst. Folglich existiert kein Beleg für eine so starke Zunahme, obwohl auch die Polizeigewerkschaft von einer Zunahme ausgeht.
Quelle:https://www.journalistenwatch.com/2017/12/22/zehn-messer-metzeleien-pro-tag-gatestone-institut-macht-merkel-dafuer-verantwortlich/(Stand: 24.6.2019).
Zusammenfassung Debunking: https://www.zeit.de/gesellschaft/zeitgeschehen/2018-09/jugendkriminalitaet-gewalt-messer-einsatz-zahlen-statistik (Stand: 24.6.2019).

dass die Inhalte vieler Fake News auch ganz explizit auf eine Polarisierung der Gesellschaft hinwirken, indem argumentativ ein Interessengegensatz zwischen Volk und Elite oder Einheimischen und Zugezogenen bekräftigt wird (vgl. Tabelle 2.12).

Betrachtet man die Themen, mit denen sich die populistischen Fake News unterschiedlichen Typs befassen, so zeigen sich interessante Zusammenhänge: Überraschend ist beispielsweise, dass in Texten, die sich mit dem Thema Migration befassen, ein noch größeres Maß an Anti-Eliten-Populismus zu finden ist als an ausgrenzendem Populismus. Auch in Texten, in denen das Thema Sicherheit im Mittelpunkt steht, werden Eliten sehr oft kritisiert. Das stützt die bereits vorgenommene Deutung: Das Migrationsthema ist in deutschsprachigen Fake News nicht nur ein Feld, in dem xenophobe Haltungen ausgelebt werden, sondern es wird sehr häufig genutzt, um etablierte Eliten zu diskreditieren (vgl. Tabelle 2.13).

Nachdem festgestellt werden konnte, dass Populismus ein integraler Bestandteil von Fake News ist, soll nun untersucht werden, wie Art und Ausmaß des Populismus mit der Menge und dem Anteil von falschen Tatsachen-

Tabelle 2.13: Arten von Populismus in Fake News zu bestimmten Themen

	Anzahl/Anteil in populistischen Fake News		Thin ⟹ Thick Art des Populismus (in %)				
Thema	**N**	**%**	**Leer**	**Minder-heiten**	**Anti-Eliten**	**Voll-ständig**	**Gesamt**
Migration und Sicherheit	134	36,5	9,0	43,3	11,9	35,8	100
Migration	53	14,4	11,3	30,2	34,0	24,5	100
Sicherheit	45	12,3	13,3	26,7	46,7	13,3	100
Justiz	34	9,3	17,6	5,9	41,2	35,3	100
Medien	24	6,5	12,5	0,0	66,7	20,8	100
Internationale Beziehungen	20	5,4	20,0	10,0	60,0	10,0	100
Sozialpolitik	15	4,1	6,7	20,0	66,7	6,7	100
Sonstige Themen	42	11,4	50,0	0,0	50,0	0,0	100
Gesamt	**367**	**100**	**16,1**	**25,3**	**34,9**	**23,7**	**100**
Gesamtanzahl/-anteil Schwerpunkt Migration und Sicherheit	**32**	**63,2**	**40,7**	**92,5**	**43,0**	**77,0**	**100**

Basis: N=367. X^2=143,71, df=21, p<0,001; Cramer's V=0,361, p<0,001. Sonstige Themen beinhaltet Arbeitsmarkt-, Wirtschafts-, Bildungs-, Finanz-, Kultur- und Europapolitik. Quelle: DORIAN-Sample, eigene Berechnung.

behauptungen korrelieren. Tabelle 2.14 zeigt, dass Meldungen mit vielen Falschinformationen auch besonders häufig populistisch sind. Es besteht ein starker Zusammenhang zwischen den Merkmalen „Anzahl der Fakes" und populistischen Argumentationsmustern (V=0,227), der hoch signifikant ist (p<0,001). Der Zusammenhang zwischen Populismus und dem Anteil der Falschaussagen am Gesamttext ist fast gleichermaßen hoch (V=0,217, p<0,001). Kurz gesagt: In populistischen Texten wird besonders häufig und auch besonders viel gelogen.

Dieser Zusammenhang ist auf zwei Arten interpretierbar: Es ist denkbar, dass populistische Argumentationsmuster genutzt werden, um Aussagen, die nicht auf Fakten basieren, zu legitimieren. Der Verweis auf einen glorifizierten Volkswillen oder auch einfach auf den gesunden „Menschenverstand",

Tabelle 2.14: Zusammenhang zwischen Fake-Anzahl, Fake-Anteil und Populismus

	Unpopulistische Fake News		Populistische Fake News		
Anzahl der Fakes	**Anzahl Fälle**	**Anteil in %**	**Anzahl Fälle**	**Anteil in %**	**Gesamt in %**
1 Fake	65	38,0	107	62,0	100
2 Fakes	32	20,6	123	79,4	100
3 Fakes	25	15,3	138	84,7	100
Gesamt	**122**	**24,9**	**368**	**75,1**	**100**
Anteil an Fakes					
< 25 %	81	34,5	155	65,5	100
< 50 %	22	14,8	127	85,2	100
> 50 %	19	18,8	82	81,1	100
Gesamt	**122**	**24,9**	**368**	**75,1**	**100**

Basis: N=489 Fake News. *Anzahl der Fakes*: X^2=25,159, df=2, p<0,001; Cramer's V=0,227, p<0,001. *Anteil der Fakes*: X^2=22,987, df=3, p<0,001; Cramer's V=0,217, p<0,001. Nicht eindeutig zuordenbare Fälle bei Anteil an Fakes (N=4) nicht ausgewiesen. Quelle: DORIAN-Sample, eigene Berechnung.

der dem Volk zugebilligt und Politikerinnen und Politikern in Abrede gestellt wird, kann dabei Tatsachenbehauptungen untermauern, die einer rationalen Überprüfung nicht standhalten würden. Auch die Anti-Eliten-Stoßrichtung vieler populistischer Aussagen kann dazu beitragen, Falschaussagen zu legitimieren: Wenn Akteure aus Politik, Journalismus und Wissenschaft als eine korrupte, eng verwobene und gegen das Volk verschworene Elite beschrieben werden, wird damit auch der diskursive Wahrheitsfindungsprozess zwischen Wissenschaft, Journalismus und Politik in Frage gestellt, auf den sich Eliten in demokratischen Systemen berufen (vgl. Waisbord, 2018).

Andererseits kann ein populistisches Weltbild dazu beitragen, dass klassische Wege der Wahrheitsfindung für überflüssig befunden werden, weil sich die Wahrheit aus dem Volkswillen speist und populistische Medienakteure sich selbst als Stimme des Volkes verstehen. Klassische journalistische Rechercheroutinen wie das Hören beider Seiten und die Notwendigkeit, sich auf zwei voneinander unabhängige Quellen zu berufen, sind mit diesem Weltbild nicht kompatibel. Wer davon ausgeht, dass er die Wahrheit dadurch erkennt, dass er auf der richtigen Seite steht, der wendet keine weiteren Qualitätssicherungsmaßnahmen an und berichtet deshalb oft falsch (Abb. 2.4).

SKB NEWS

EIN MAINSTREAMUNABHÄNGIGER NACHRICHTEN -UND INFORMATIONSKANAL – HERAUSGEBERIN: GABY KRAAL

DONNERSTAG, AUGUST 16TH, 2018

Search...

MENÜ

GESELLSCHAFT

Folgen

Degenerierte Politikergeneration: Wenn Kriminelle besser versorgt werden, als unsere Schulkinder

BY GABY KRAAL ON 14. DEZEMBER 2017 • (31 KOMMENTARE)

Justizsenator Dirk Behrendt von den Grünen/Bündnis90 in Berlin ist der Meinung: „Das sei wichtig für die Resozialisierung"

Den Häftlingen sollte der Aufenthalt im Gefängnis einst eine Lehre sein, ...aber das war zu einer Zeit, bevor inkompetente und realitätsferne Politiker die Macht in Deutschland übernahmen.

Viele Justizvollzugsanstalten in Deutschland sehen heute besser aus, als die Schulen, in der unsere Kinder lernen sollen, um anschließend mit vernünftiger Bildung und guten Jobaussichten ein ehrliches und anständiges Leben führen zu können. – Aber:

Die heutigen Haftanstalten, wir schauen einmal nach Berlin, sind besser ausgestattet als so manche Schule.

Die Justizvollzugsanstalt Heidering in Großbeeren ist so ein Beispiel. Hier sitzen bis zu 647 Häftlinge ein. Sie sollen einen kostenlosen Internetzugang und mobile Tabletcomputer bekommen. Der Berliner

Abbildung 2.4: Populistische Fake News mit starkem Gegensatz Volk-Elite

Politiker werden als „realitätsfern“ und „degeneriert“ dargestellt. Es wird betont, dass sie deshalb lieber Kriminelle als Schulkinder versorgen, die in diesem Artikel für das Volk stehen. Im weiteren Artikel werden dann aus ganz Deutschland Beispiele für eine – laut Autorin – zu gute Versorgung von Strafgefangenen angeführt.
Die meisten Beispiele haben zwar einen realen Hintergrund, wichtige Details wurden aber hinzuerfunden oder weggelassen. So erhalten zwar einige ausgesuchte Häftlinge in Berlin Internetzugang, aber Zugriff haben sie ausschließlich auf Lernsoftware, Wikipedia, Stellenangebote und Nachrichtenportale.
Häufig bei Fake News anzutreffen ist auch die hier verwandte Textstruktur: Am Anfang der Meldung steht eine Interpretation/Schlussfolgerung und nicht der Anlass zu Bericht bzw. das Aktuellste wie im klassischen Journalismus (vgl. Abschnitt C.)
Quelle:
https://web.archive.org/web/20180314154506/https://schluesselkindblog.com/2017/12/14/degenerierte-politikergeneration-wenn-kriminelle-besser-versorgt-werden-als-unsere-schulkinder/ (Stand: 24.6.2019).

Populismus kann also als eine Strategie verstanden werden, um falsche Tatsachenbehauptungen zu verschleiern, oder auch als eine Ursache dafür, dass Falschbehauptungen entstehen. Gleichgültig, welchen kausalen Zusammenhang man als plausibel erachtet: Der Populismus politisch-medialer Akteure und ihr Hang, Falschinformationen zu verbreiten, können als einander verstärkende Phänomene gedeutet werden.

Ein weiterer interessanter Zusammenhang zwischen populistischen Inhalten und Fake News wird in Tabelle 2.15 deutlich: Populistische Fake News platzieren häufiger als nicht-populistische Fake News die Lüge bereits in Überschrift und Teaser, also in jenen Elementen des Beitrags, die die Aufmerksamkeit der Rezipierenden erregen sollen. Sie folgen somit in besonders starkem Maße dem Muster, das bereits bei der Beschreibung des Gesamtsamples als Anpassung an die Logik von Social Networks gedeutet wurde: Die falsche Tatsachenbehauptung in Überschrift, Bild oder Teaser lockt dabei Leserinnen und Leser an. Auffällig ist allerdings, dass populistische Texte sehr viel häufiger als nicht-populistische Texte die falsche Tatsachenbehauptung auch im Text wiederholen.

Fake-News-Produzierende können also als politisch-mediale Akteure verstanden werden, die Meldungen verfassen, in denen sehr häufig populistische Argumentationsmuster zu finden sind. Für die Verbreitung dieser Meldungen

Tabelle 2.15: Position der Falschinformation in Fake News

Position der falschen Tatsachenbehauptung	Unpopulistische Fake News (N=122)	Populistische Fake News (N=367)	Gesamt
Bild	45,5	54,5	100
Überschrift/Teaser	27,3	72,7	100
Text	32,4	67,6	100
Überschrift/Teaser und Text	12,8	87,2	100
Gesamt	**24,9**	**75,1**	**100**

Basis: N=489 Fake News. Angaben in %. *Position der falschen Behauptung*: X^2=18,311, df=3, p<0,001; Cramers V=0,194, p<0,001. Quelle: DORIAN-Sample, eigene Berechnung.

hoffen sie dabei unter anderem auf die Social-Network-Accounts von populistischen Politikerinnen und Politikern, die sie als natürliche Verbündete verstehen.[21] Aber wie gut funktioniert dieser Verbreitungsweg? Welche Verbreitungswege bestehen darüber hinaus und wie könnten diese – aufbauend auf die hier generierten Erkenntnisse über Strukturen und Strategien in Fake News – weiter untersucht werden?

F. Fazit und Ausblick: Wie wirkt die Struktur auf die Verbreitung? - Weiterer Forschungsbedarf

In diesem Kapitel lag der Fokus darauf, zu erfassen, welche strukturellen, sprachlichen, argumentativen und thematischen Muster Fake News im deutschsprachigen Raum aufweisen. Dafür wurde ein Sample von knapp 500 im Internet verbreiteten Beiträgen untersucht, die alle mindestens eine falsche Tatsachenbehauptung enthalten. Die vorliegenden Analysen haben dazu beigetragen zu verstehen, **was** verbreitet wird. Die Frage **wie** die Verbreitung genau vonstattengeht, wurde zurückgestellt, obwohl sie von gesellschaftlich großer Bedeutung ist, denn erst durch ihre virale Verbreitung in Social Networks können Fake News Wirkung entfalten.

21 https://philosophia-perennis.com/2019/05/12/der-erste-kongress-der-freien-medien-im-bundestag-eine-sternstunde-des-journalismus/(Stand: 24.6.2019), https://juergenfritz.com/2019/05/12/afd-puscht-mainstreammedien/ (Stand: 24.6.2019).

Der Entscheidung, so vorzugehen, liegt folgende Überlegung zugrunde: Sowohl auf den deutschsprachigen Raum bezogen als auch international existieren bereits aktuelle und empirisch fundierte Studien, die sich mit den Verbreitungswegen von Fake News befassen (Benkler et al., 2018; Marchal et al., 2019; Neudert et al., 2017; Sängerlaub et al., 2018; Vosoughi et al., 2018). Da diese Studien jedoch in Bezug auf die Beschaffenheit von Fake News nur ungefähre Erkenntnisse enthalten, können in ihnen kaum Aussagen darüber getroffen werden, welche Eigenschaften von Fake News eine Verbreitung begünstigen oder erschweren. Zu wissen, welche Fake News das größte Potenzial haben, sich viral zu verbreiten, kann für ihre Bekämpfung jedoch von entscheidender Bedeutung sein. Die Vorgehensweise in der hier vorgestellten Studie basiert also auf der Überzeugung, dass auf Grundlage von umfassenden Erkenntnissen über die Beschaffenheit von Fake News in einem zweiten Schritt und unter Fortführung der interdisziplinären Zusammenarbeit präzisere Aussagen über ihre Viralität getroffen und wirkungsmächtige Mechanismen zur Fake-News-Erkennung und -Bekämpfung entwickelt werden können.[22]

Im Folgenden werden deshalb nicht nur die hier dargestellten Erkenntnisse über Textstruktur, Journalismus-Imitation, thematische Ausrichtung und Argumentationsstrukturen von Fake News zusammengefasst, sondern auch in anderen Studien gewonnene Erkenntnisse über die Verbreitung von deutschsprachigen Fake News wiedergegeben und daraus eine Agenda für weitergehende Forschung generiert.

Besonders einschlägig für den Forschungsstand zur Fake-News-Verbreitung im deutschsprachigen Raum ist dabei die bereits mehrfach zitierte Studie von Sängerlaub et al., die im direkten Umfeld der Bundestagswahl 2017 die Verbreitung von zehn Falschmeldungen auf Social Networks und Nachrichtenseiten untersucht hat, wobei parallel dazu auch ausgewertet wurde, wie sich von Behörden, Fakt-Checking-Akteuren und journalistischen Akteuren erstellte Debunking-Nachrichten verbreiten (Sängerlaub et al., 2018: 2).

Kernergebnis dieser Studie zur Verbreitung von Fake News ist, dass im deutschsprachigen Internet wenige professionelle Kanäle, die von Politikerinnen und Politikern, Publizisten und Aktivisten betrieben werden, für den

22 Auch die Untersuchung der Verbreitungswege bietet Ansätze zur Fake-News-Erkennung und -Bekämpfung, beispielsweise, wenn Meldungen von Facebook-Seiten, die erst kurze Zeit bestehen, im Newsfeed automatisch schlechter gestellt werden (vgl. Weedon et al., 2017). Die Wirksamkeit solcher Verfahren wird hier nicht verneint, sondern soll vielmehr ergänzt werden.

Löwenanteil der Verbreitung von Fake News zuständig sind: So wurden von Sängerlaub et al. für jeden falschen Bericht zehn Top-Verbreiter identifiziert und dann errechnet, dass diese zehn Accounts durchschnittlich für 56 Prozent des Gesamt-Engagements zuständig sind (Sängerlaub et al., 2018: 85).[23] Dies bedeutet, dass für die Verbreitung von Fake News im deutschsprachigen Internet nicht nur klassische Viralität im Sinne des Teilens von zahlreichen einzelnen Nutzenden eine Rolle spielt, sondern die Entscheidung einzelner Akteure mit hoher Reichweite von großer Bedeutung ist. Als zentral für die Verbreitung von Fake News erwies sich in der Studie die AfD, die in sieben von zehn untersuchten Fällen zu den zehn Topverbreitern gehörte. Dies zeigt, dass die Aktivisten der sogenannten „alternativen Medien", die AfD-Politikerinnen und -Politiker zum verstärkten Teilen ihrer Beiträge aufrufen (vgl. Abschnitt E.), absolut rational handeln, denn Social-Network-Kanäle der Partei und einzelner Politiker tragen wesentlich zur Verbreitung von Fake-News-Inhalten bei.

Ebenfalls relevante Ergebnisse zur Verbreitung von Desinformation im deutschsprachigen Raum liefern Studien des „Oxford Internet Institute", das sowohl im direkten Vorfeld der Bundestagswahl 2017 als auch vor der Europawahl 2019 untersucht hat, welcher Anteil der auf Twitter (Neudert et al., 2017) bzw. auf Twitter und Facebook (Marchal et al. 2019) geteilten Nachrichten von seriösen Seiten stammt und welcher Anteil von Junk-Seiten[24], die dafür bekannt sind, auch Fake News zu verbreiten (Marchal et al., 2019; Neudert et al., 2017).[25] Beide Studien zeigen, dass in Deutschland auf Twitter sehr viel mehr seriöse Nachrichten als Beiträge von Junk-Seiten verbreitet werden (Marchal et al., 2019: 3; Neudert et al., 2017). Die Studie zur Europawahl zeigt allerdings auch, dass auf Facebook Junk-Nachrichten zwar in ihrer Reichweite nicht an seriöse Nachrichten heranreichen, dafür aber oft ein

23 Einschränkend lässt sich anmerken, dass Sängerlaub et al. zwar feststellten, dass Facebook für die Verbreitung von Fake News in Deutschland der wichtigste Kanal ist, dass sie allerdings gerade bei Facebook nur öffentliche Seiten untersuchen konnten (2018: 2). Öffentliche Seiten gehören wiederum eher politischen und politisch-medialen Akteuren, während private Seiten, private Gruppen und die Weiterleitung über Kanäle wie Whats-App für die Viralität von Artikeln eine große, in der Studie unerkannte Rolle spielen könnten.

24 Als Junk-Seiten verstehen die Autoren Webseiten, die ideologisch extreme, irreführende und sachlich falsche Informationen verbreiten (Marchal et al., 2019: 1).

25 Die Studie basiert auf einer sehr großen Datenmenge, Fakt-Checking von einzelnen Beiträgen fand deshalb nicht statt. Hier wurden also keine Fake News an sich untersucht, sondern allgemeiner Nachrichten, die von verdächtigen Seiten stammen.

besonders hohes Engagement im Sinne von teilen, liken und kommentieren auslösen (Marchal et al., 2019).

I. Struktur, Fake-Anteil und Verbreitung

Die Untersuchung des Fake-News-Samples hat gezeigt: Bei Fake News im deutschsprachigen Raum handelt es sich in den meisten Fällen nicht um durchweg erfundene Berichte, sondern um Beiträge, in denen reale und verfälschte Fakten gemischt werden, mitunter noch durch frei erfundene Details ergänzt (vgl. Tabelle 2.2): Fast die Hälfte der Artikel enthält weniger als 25 Prozent falsche Behauptungen. Nur jeder fünfte Artikel in unserem Sample bestand zu mehr als 50 Prozent aus falschen Behauptungen. Zugleich wurde die falsche Tatsachenbehauptung in 57 Prozent der Fälle auch oder ausschließlich in Überschrift oder Teaser platziert, also in jenen Elementen des Artikels, die im Internet dazu beitragen, Aufmerksamkeit zu erregen.

Wie sich die jeweils spezifische Mischung von richtigen und falschen Fakten und die Platzierung der falschen Tatsachenbehauptungen auf die Verbreitung der Fake News auswirkt, wurde bisher nur ansatzweise untersucht.

Für den deutschsprachigen Raum haben Sängerlaub et al. (2018: 70) anhand der zehn von ihnen untersuchten exemplarischen Fälle aufgezeigt, dass mitunter auch Meldungen, deren Berichtsgegenstand frei erfunden ist („Fabricated Content“, siehe dazu auch Wardle, 2017), relativ große Verbreitung finden. Zugleich zeigen sie aber auch, dass wichtige Multiplikatoren wie führende AfD-Politikerinnen und -Politiker eher Meldungen verbreiten, in denen nicht alles erfunden ist, aber wichtige Fakten falsch interpretiert oder manipuliert wurden („Misinterpreted“ und „Manipulated Content“, siehe dazu auch Wardle, 2017). Typische Verbreitungswege von Fake News mit jeweils unterschiedlichem Wahrheitsgehalt und unterschiedlicher Platzierung der Falschinformation auf Grundlage einer größeren Fallzahl nachzuvollziehen, wäre hierbei von großem Interesse.

US-Studien zur Verbreitung von Inhalten auf Twitter haben gezeigt, dass auf Twitter falsche Neuigkeiten mit einer um 70 Prozent höheren Wahrscheinlichkeit verbreitet werden als wahre Inhalte und dass viele Menschen Meldungen teilen, von denen sie nur die Überschrift gelesen haben (Gabielkov et al., 2016; Vosoughi et al., 2018). Wenn diese im englischen Sprachraum generierten Erkenntnisse auf den deutschsprachigen Raum übertragen werden können, ist die Taktik deutscher Fake-News-Produzierenden, die Lüge bereits in der Überschrift zu platzieren, der Verbreitung von Fake News in

hohem Maße dienlich. In eine ähnliche Richtung deuten die Erkenntnisse von Marchal et al. (2019), die in einer europaweiten Studie zeigen konnten, dass Beiträge von Junk-Seiten auf Facebook mehr Engagement auslösen als Beiträge seriöser Medien.

Gleichzeitig ergibt sich aus der hier beschriebenen Struktur von Fake News auch ein Ansatz zur Bekämpfung: Wenn über die Hälfte der Fake News bereits in Überschrift und Teaser eine falsche Tatsachenbehauptung enthalten und genau diese Meldungen besonders hohes Viralitätspotenzial bieten (was für den deutschsprachigen Raum allerdings noch nachgewiesen werden müsste), können schon Maßnahmen, die sich auf diese Textteile konzentrieren, wesentlich dazu beitragen, Fake News einzudämmen.

II. Professionalitätsgrad, Nachrichtenfaktoren und Verbreitung

Hinsichtlich des Professionalitätsgrads zeigen die Analysen, dass es den Fake-News-Produzierenden in Deutschland überwiegend (aber keinesfalls immer) gelingt, ein journalistisches Erscheinungsbild zu imitieren und grobe orthographische und grammatikalische Fehler zu vermeiden (vgl. Tabelle 2.4). Was Textaufbau und interne Konsistenz der Fakten anbelangt, genügen die meisten Fake News jedoch nicht einmal annähernd den Qualitätsansprüchen von professionellem Journalismus (vgl. Tabelle 2.5). Dabei fällt auf, dass die von Rezipienten zuerst wahrgenommenen und für das Teilen von Artikeln besonders relevanten Elemente Überschrift und Leadsatz (vgl. Gabielkov et al. 2016) in Sachen Professionalität noch besser abschneiden als die eigentlichen Texte.

Dies bedeutet, dass für Menschen, die mit journalistischen Qualitätsmaßstäben gut vertraut sind, ein großer Teil der deutschsprachigen Fake News derzeit noch relativ leicht zu erkennen ist – zumindest sofern sie bereit sind, die Meldungen sorgfältig und bis zum Ende zu lesen und dabei auf Konsistenz der Fakten, Textaufbau und den nicht-journalistischen Umgang mit Mutmaßungen zu achten. Für die Bekämpfung von Fake News folgt daraus: Medienbildung und Aufklärung bezüglich der Qualitätsmaßstäbe des seriösen Journalismus versetzen Internet-Nutzende zugleich immer auch in die Lage, Fake News zu erkennen. Journalismus sollte sich deshalb bemühen, nicht nur unterscheidbar zu bleiben, sondern seine eigenen Qualitätsmaßstäbe offensiv zu kommunizieren. In Folgestudien könnte untersucht werden, ob Fake News mit einem höheren Professionalitätsgrad sich auch schneller verbreiten, was soweit keinesfalls nachgewiesen ist. So zeigen beispielsweise Sängerlaub et

al. (2018: 41), dass sich auch plump gefälschte Polizei-Dienstanweisungen voller Rechtschreibfehler viral verbreiten können.

Bei der Untersuchung von Nachrichtenfaktoren wurde zudem deutlich, dass sich Fake News im deutschsprachigen Raum in hohem Maße an Nachrichtenfaktoren orientieren, wobei Nähe, Negativität und Etablierung besonders stark berücksichtigt werden (vgl. Tabelle 2.7). Interessant wäre hier eine Folgeuntersuchung, die den Einfluss von bestimmten Nachrichtenfaktoren auf die Verbreitung von Fake News zum Gegenstand hat. Anknüpfend an die Erkenntnis, dass einzelne professionell geführte Accounts von politischen und politisch-medialen Akteuren für einen guten Teil der Reichweite von Fake News im deutschsprachigen Raum zuständig sind (Sängerlaub et al., 2018: 85), könnte zudem untersucht werden, ob diese Akteure an Nachrichtenfaktoren orientiertes Gate-Keeping betreiben, um die Aufmerksamkeit ihrer Nutzerinnen und Nutzer zu optimieren, und welche Nachrichtenfaktoren die Auswahl dieser Akteure zentral bestimmen. So könnte das politisch-mediale Zusammenspiel im rechtspopulistischen Milieu besser verstanden werden.

III. Thematische Ausrichtung und Verbreitung

Im Sample dominieren Fake News, die sich mit den Themen Kriminalität, Migration oder einer Kombination aus beiden Themen befassen: Sie machen beinahe zwei Drittel der Falschmeldungen aus (vgl. Tabelle 2.9).

Untersucht man die Nachrichtenfaktoren bei Meldungen, die sich mit den Themen Migration und innere Sicherheit befassen, gesondert, wird eine Konzentration auf die Nachrichtenfaktoren Nähe und Schaden deutlich (Tabelle 2.10). Fake News in Deutschland erzählen also sehr oft Geschichten über angeblich oder tatsächlich kriminelle Flüchtlinge oder Migranten und sie erzählen diese Geschichten auf eine Art und Weise, die die Rezipienten dazu einlädt, die Verbrechen als das eigene Umfeld bedrohend wahrzunehmen. Auffällig ist zudem, dass gerade in der Berichterstattung zum genannten Themenfeld besonders häufig korrekte und erfundene Fakten vermischt werden (vgl. Tabelle 2.11). Vorfälle werden also meist nicht frei erfunden, sondern reale Vorfälle überzeichnet, Details hinzuerfunden und aus einer Aneinanderreihung von Beispielen statistische Aussagen generiert, die nicht belegt werden. Dies macht es für Laien besonders schwierig, Fake News zum Thema Migration und innere Sicherheit als solche zu erkennen, denn selbst wenn Rezipierende die Übertreibung vermuten, bleibt das Gefühl, dass ein wahrer

Kern besteht – ein Gefühl, das durch die Betonung von Nähe Angst auslösen kann.

Über den Zusammenhang von thematischer Ausrichtung und Verbreitung von Fake News ist derzeit wenig bekannt. Genau wie im vorliegenden Sample wurden auch bei den Untersuchungen von Humprecht (2018) und Sängerlaub et al. (2018) Fake News aus dem Themenfeld Migration und innere Sicherheit als charakteristisch für die deutschsprachige Fake-News-Szene ausfindig gemacht. Sängerlaub und Kollegen sprechen in diesem Kontext von einem „Angstnarrativ", das zur Verbreitung von Fake News weit über das rechtspopulistische Milieu hinaus beitrage und das in einzelnen Fällen auch von CSU-Politikern bedient worden sei (2018: 64).

Weitere Forschung könnte hier themenspezifische Entstehungs- und Verbreitungswege von Fake News offenlegen. Viele Fake News zum Thema Migration und innere Sicherheit basieren direkt (oder indirekt auf dem Umweg über Lokalberichterstattung) auf Polizeimeldungen und werden dann von zahlreichen Webseiten mit Fake-News-Anteil aufgegriffen und weiterverbreitet. Genauere Erkenntnisse über solche themenspezifischen Verbreitungswege wären deshalb auch hilfreich, um die Pressearbeit der Polizei im digitalen Raum weiter zu professionalisieren. Zudem müsste untersucht werden, wie es populistischen Medienakteuren in Zusammenarbeit mit politischen Akteuren gelingt, Nachrichten über Migration und Kriminalität, über die zunächst nur lokal berichtet wurde, in der nationalen Medienagenda zu etablieren.

IV. Populismus und Verbreitung

Ein Zusammenhang zwischen Populismus und Fake News wurde zwar bereits oftmals vermutet, konnte im Rahmen der vorliegenden Untersuchung jedoch empirisch bewiesen werden: Ein großer Teil der Fake News im deutschsprachigen Raum enthält populistische Argumentationsmuster (vgl. Tabelle 2.12). Darüber hinaus beinhalten populistische Fake News oftmals einen besonders hohen Anteil von falschen Tatsachenbehauptungen. Meist wird dabei nicht nur das Volk oder der Volkswillen glorifiziert, sondern es werden zusätzlich noch Minderheiten oder Eliten diffamiert (vgl. Tabelle 2.13).

Auffällig ist, dass deutschsprachige Fake News trotz ihrer thematischen Fixierung auf Migration und Kriminalität noch häufiger eine antielitäre Stoßrichtung enthalten als Kommunikationsstrategien, die sich gegen Minderheiten richten (vgl. Tabelle 2.14). Dies ist sogar erkennbar, wenn man die Gruppe von Fake News, die sich thematisch mit Migration und innerer Si-

cherheit befassen, gesondert betrachtet. Das Angstnarrativ (Sängerlaub et al., 2018: 64) ist also mitnichten unpolitisch, sondern oftmals schlussendlich auf die Diffamierung von Eliten ausgerichtet.

Die Bedeutung von populistischen Politikerinnen und Politikern und deren Social-Media-Accounts für die Verbreitung von Fake News im deutschsprachigen Raum zeigen bereits Sängerlaub et al. (2018) auf. Noch nicht untersucht wurde allerdings, inwieweit die populistischen Argumentationsmuster selbst zur Verbreitung von Fake News beitragen: Werden Fake News mit explizit populistischem Narrativ auch besonders häufig von populistischen Politikerinnen und Politikern geteilt? Führen populistische Kommunikationsstrategien bei Nutzerinnen und Nutzern mit geringerer Reichweite eher dazu, dass sie Fake News weiterverbreiten oder wirken sie eher abschreckend? Dies könnte in weiteren Studien geklärt werden.

Um die Verbreitung von Fake News basierend auf den hier generierten Erkenntnissen zur Beschaffenheit von Fake News zu untersuchen, ist interdisziplinäre Zusammenarbeit unbedingt notwendig (vgl. auch Vosoughi et al., 2018: 1150). Nur bei einer eng verzahnten Zusammenarbeit von unterschiedlichen Disziplinen kann verfolgt werden, welche Eigenschaften von Fake News zur Verbreitung beitragen (Kommunikationswissenschaft), auf welchem Wege die Verbreitung erfolgt (Informatik) und welche psychologischen Mechanismen Nutzerinnen und Nutzer dazu bewegen, Fake News weiter zu teilen oder aufgrund von Fake News ihre Einstellung zu ändern (Medienpsychologie). Die Bekämpfung von Fake News muss wiederum auf einer entsprechend informierten rechtlichen Rahmung basieren (Recht). Weil Fake News ein Phänomen darstellen, das sich kontinuierlich verändert und das zugleich das Potenzial birgt, die demokratischen politischen Kommunikationsprozesse nachhaltig zu beeinflussen, ist kontinuierliche und im Sinne einer Eindämmung anwendungsorientierte Forschung dringend geboten.

Literaturverzeichnis zu Kapitel 2

Aalberg, T., Esser, F., & Reinemann, C. (2017). Populist political communication in Europe. Routledge research in communication studies. New York: Routledge.

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211-236. https://doi.org/10.1257/jep.31.2.211

Bednarek, M., & Caple, H. (2017). The Discourse of News Values: How News Organizations Create ‚Newsworthiness'. New York: Oxford University Press.

Benkler, Y., Faris, R., & Roberts, H. (2018). Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. New York, NY: Oxford University Press.

Blassnig, S., Ernst, N., Büchel, F., Engesser, S., & Esser, F. (2018). Populism in Online Election Coverage. Analyzing Populist Statements by Politicians, Journalists and Readers in Three Countries. Journalism Studies, 38(1), 1-20. https://doi.org/10.1080/1461670X.2018.1487802

Brinkschulte, F., Klapproth, J. J., & Frischlich, L. (11.05.2019). ‚Wenn Sie die Wahrheit zu schätzen wissen': Die Integration von Mainstram-Medien und alternativen Pseudo-Presse-Angeboten als Quellen rechtspopulistischer Medienmacher. Vortrag auf der 64. Jahrestagung der Deutschen Gesellschaft für Publizistik und Kommunikationswissenschaft (DGPuK), Münster.

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. Annals of the International Communication Association, 43(2), 97-116. https://doi.org/10.1080/23808985.2019.1602782

Engesser, S., Ernst, N., Esser, F., & Büchel, F. (2017). Populism and Social Media: How Politicians Spread a Fragmented Ideology. Information, Communication & Society 20(8), 1109–1126. https://doi.org/10.1080/1369118X.2016.1207697

Engesser, S., Fawzi, N., & Larsson, A. O. (2017). Populist Online Communication: Introduction to the Special Issue. Information, Communication & Society, 20(9), 1279–1292. https://doi.org/10.1080/1369118X.2017.1328525

Flath, H. (2013). Storytelling im Journalismus. Formen und Wirkungen narrativer Berichterstattung (Dissertation). Universität Ilmenau. Verfügbar unter: https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00027890/ilm1-2013000242.pdf

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016). Social Clicks: What and Who Gets Read on Twitter? ACM SIGMETRICS Performance Evaluation Review, 44(1), 179–192. https://doi.org/10.1145/2896377.2901462

Galtung, J., & Ruge, M. H. (1965). The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers. Journal of Peace Research, 2(1), 64–91. https://doi.org/10.1177/002234336500200104

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1), eaau4586 (1-8). https://doi.org/10.1126/sciadv.aau4586

Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2018). Avoiding the Echo Chamber About Echo Chambers: Why Selective Exposure to Like-Minded Political News is Less Prevalent Than You Think. Miami, FL: John S. & James L. Knight Foundation.

Hindman, M., & Barash, V. (2018). Disinformation, ‚Fake News' and Influence Campaigns on Twitter. Miami, FL. Verfügbar unter: https://kf.org/2IGNOQV

Holt, K., Figenschou, T. U., & Frischlich, L. (2019). Key Dimensions of Alternative Media. Digital Journalism (Online). https://doi.org/10.1080/21670811.2019.1625715

Hooffacker, G., & Meier, K. (2017). La Roches Einführung in den praktischen Journalismus: Mit genauer Beschreibung aller Ausbildungswege Deutschland - Österreich - Schweiz (20., neu bearbeitete Aufl.). Journalistische Praxis. Wiesbaden: Springer VS. http://dx.doi.org/10.1007/978-3-658-16658-8

Horstmann, A. C., Rösner, L., Conrad, L., & Heidemann, R. (2018, September). Fake news or real truth?! Ergebnisse einer Think Aloud Befragung zur Erkennung von Falschnachrichten. 51. Kongress der Deutschen Gesellschaft für Psychologie (DGP), Frankfurt.

Humprecht, E. (2018). Where 'Fake News' Flourishes: A Comparison Across Four Western Democracies. Information, Communication & Society: Online first, 1–16. https://doi.org/10.1080/1369118X.2018.1474241

Jagers, J., & Walgrave, S. (2007). Populism as political communication style: An empirical study of political parties' discourse in Belgium. European Journal of Political Research, 46(3), 319–345. https://doi.org/10.1111/j.1475-6765.2006.00690.x

Kepplinger, H. M. (2008). Reciprocal Effects. In The international encyclopedia of communication: Precision journalism - Rhetoric in Western Europe: Britian. (IX, S. 4143–4147). Malden, MA, Oxford, Carlton, Victoria: Blackwell Publishing Ltd.

Klinger, U., & Svensson, J. (2015). The emergence of network media logic in political communication: A theoretical approach. New Media & Society, 17(8), 1241–1257. https://doi.org/10.1177/1461444814522952

LaRoche, W. v., Hooffacker, G., & Meier, K. (2013). Einführung in den praktischen Journalismus: Mit genauer Beschreibung aller Ausbildungswege Deutschland Österreich Schweiz (19., überarb. u. aktualisierte Aufl.). Journalistische Praxis. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The Science of Fake News. Science (New York, N.Y.), 359(6380), 1094–1096, https://doi.org/10.1126/science.aao2998

Lee, J. H. (2009). News Values, Media Coverage, and Audience Attention: An Analysis of Direct and Mediated Causal Relationships. Journalism & Mass Communication Quarterly 86(1), 175–190. https://doi.org/10.1177/107769900908600111

Levy, N. (2017). The bad news about fake news. Social Epistemology Reviews and Reply Collective – Online, 6(8), 20-36. Zugriff am 26.11.2018. Verfügbar unter http://wp.me/p1Bfg0-3GV

Liesem, K. (2015). Professionelles Schreiben für den Journalismus. Wiesbaden: Springer Fachmedien Wiesbaden. Verfügbar unter: http://dx.doi.org/10.1007/978-3-531-19008-2

Mahatpatra, S., & Plagemann, J. (2019). Polarisation and Politicisation: The Social Media Strategies of Indian Political Parties (GIGA Focus | ASIA No. 3). Hamburg. Verfügbar unter: https://www.giga-hamburg.de/de/system/files/publications/gf_asien_1903_en.pdf

Marchal, N., Kollanyi, B., Neudert, L.-M., & Howard, P. N. (2019). Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook. (Data Memo No. 2019.3). Oxford, UK. Verfügbar unter: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp_GermanElections_Sep2017v5.pdf

Marwick, A., & Lewis, R. (2017). Media Manipulation and Disinformation Online. Verfügbar unter: https://datasociety.net/output/media-manipulation-and-disinfo-online/

Mudde, C. (2004). The Populist Zeitgeist. Government and Opposition 39(4), 542–563. https://doi.org/10.1111/j.1477-7053.2004.00135.x

Mudde, C., & Rovira Kaltwasser, C. R. (2016). Populism: A Very Short Introduction. Very Short Introductions: Vol. 510. New York, NY: Oxford University Press.

Nelson, J. L., & Taneja, H. (2018). The Small, Disloyal Fake News Audience: The Role of Audience Availability in Fake News Consumption. New Media & Society, 86(7), 146144481875871. https://doi.org/10.1177/1461444818758715

Neudert, L.-M., Kollanyi, B., & Howard, P. N. (2017). Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter? (Data Memo No. 2017.7). Oxford, UK. Verfügbar unter: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/ComProp_GermanElections_Sep2017v5.pdf

Plöchinger, S. (2014). Wie man erfolgreich eine Seite macht. „Blattmachen" im Netz. In C. Jakubetz (Ed.), Universalcode: Journalismus im digitalen Zeitalter (2. Aufl., S. 43–81). Affing: EFF ESS Verlagsanstalt.

Reinemann, C., Aalberg, T., Stanyer, J., Esser, F., & Vreese, C. H. de (Hrsg.). (2019). Communicating populism: Comparing actor perceptions, media coverage, and effects on citizenship in europe. New York: Routledge, Taylor & Francis Group.

Riedlinger, E., & von Detten, I. (2018). „Fake-News"-Netzwerke im deutschsprachigen Raum. Eine Struktur-Analyse anhand von Hyperlinks. Hochschule der Medien Stuttgart, Stuttgart. Unveröffentlichte Bachelorarbeit, im Druck für den Tagungsband 2019 des Düsseldorfer Forums Politische Kommunikation (DFPK).

Ruhrmann, G., & Göbbel, R. (2007). Veränderung der Nachrichtenfaktoren und Auswirkungen auf die journalistische Praxis (nr-Studie). Berlin. Verfügbar unter: https://netzwerkrecherche.org/wp-content/uploads/2015/02/nr-studie-nachrichtenfaktoren.pdf

Sängerlaub, A., Meier, M., & Rühl, W.-D. (2018). Fakten statt Fakes: Das Phänomen „Fake News". Verursacher, Verbreitungswege und Wirkungen von Fake News im Bundestagswahlkampf 2017 (Abschlussbericht Projekt „Measuring Fake News"). Berlin. Verfügbar unter: https://www.stiftung-nv.de/sites/default/files/snv_fakten_statt_fakes.pdf

Scholl, A. & Völker, J. (2019). Fake News, aktuelle Desinformationen und das Problem der Systematisierung. Anmerkungen zum Aufsatz von Fabian Zimmermann & Matthias Kohring „‚Fake News' als aktuelle Desinformation – systematische Bestimmung eines heterogenen Begriffs" in M&K 4/2018. Medien & Kommunikationswissenschaft, 67(2), 206-214.

Shin, J., Jian, L., Driscoll, K. & Bar, F. (2018). The diffusion of misinformation on social media: Temporal pattern, message, and source. Computers in Human Behavior, 83, 278-287.

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

Shearer, E., & Matsa, K. E. (2018). News Use Across Social Media Platforms 2018 (Reports 2018). Washington, US. Verfügbar unter: https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/

Tumber, H., & Waisbord, S. (2017). The Routledge Companion to Media and Human Rights. Milton: Taylor and Francis. Verfügbar unter: https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4891093

Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True and False News Online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vreese, C. H. de, Esser, F., Aalberg, T., Reinemann, C., & Stanyer, J. (2018). Populism as an Expression of Political Communication Content and Style: A New Perspective. The International Journal of Press/Politics, 23(4), 423–438. https://doi.org/10.1177/1940161218790035

Waisbord, S. (2018). Why Populism is Troubling for Democratic Communication. Communication, Culture and Critique, 11(1), 21–34. https://doi.org/10.1093/ccc/tcx005

Wardle, C. (2017). Fake News. Es ist kompliziert. Verfügbar unter: https://de.firstdraftnews.org/fake-news-es-ist-kompliziert/

Wardle, C., & Derekhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. (Council of Europe report DGI (2017)09). Strasburg. Verfügbar unter: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Warren, C. (1934). Modern news reporting. New York: Harper & Brothers.

Weedon, J., Nuland, W., & Stamos, A. (2017). Information Operations and Facebook. Verfügbar unter: https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

Wendelin, M., Engelmann, I., & Neubarth, J. (2014). Nachrichtenfaktoren und Themen in Nutzerrankings. Ein Vergleich der journalistischen Nachrichtenauswahl und der Selektionsentscheidungen des Publikums im Internet. M&K Medien & Kommunikationswissenschaft, 62(3), 439–458. https://doi.org/10.5771/1615-634x-2014-3-439

Zimmermann, F., & Kohring, M. (2018). „Fake News" als aktuelle Desinformation. Systematische Bestimmung eines heterogenen Begriffs. M&K Medien & Kommunikationswissenschaft, 66(4), 526–541. https://doi.org/10.5771/1615-634X-2018-4-526

Kapitel 3: Wirkung und Bekämpfung von Desinformation aus medienpsychologischer Sicht

Autorinnen:

Birte Högden
Prof. Dr. Nicole Krämer
Judith Meinert
Dr. Leonie Schaewitz

A. Einleitung

Medienpsychologische Forschung widmet sich schon seit mehreren Jahrzehnten den Wirkungen von Desinformation. Obwohl das Phänomen eine höhere Aufmerksamkeit erhält seit sich Falschinformationen über Social Networks verbreiten, ist es medienpsychologisch bereits seit langem relevant, zu untersuchen, wie langfristig falsche, über die Medien rezipierte Informationen im Gedächtnis behalten werden und wie sie überschrieben und korrigiert werden können. Der im Rahmen dieser Forschung identifizierte Falschinformationseffekt zeigt dabei, dass eine Korrektur oft nicht gelingt. Während diese Erkenntnisse eher Anlass zur Sorge geben, gibt es andere medienpsychologische Forschungszweige, die insgesamt aufzeigen, dass Menschen sich im digitalen Raum kompetent verhalten und beispielsweise die zugeschriebene Glaubwürdigkeit auf Basis der Kompetenz der Quelle vornehmen. Die in diesem Kapitel vorgenommene Analyse der bisherigen Literatur und Empirie sowie darauf aufbauende weitere eigene Studien sollen somit im Folgenden einen differenzierten Blick darauf ermöglichen, welche Gefahren Falschinformationen darstellen und wie diesen begegnet werden kann.

B. Überblick über die Forschungslandschaft

Medienpsychologische Forschung widmet sich traditionell vor allem der Frage, welche kurz- und langfristigen Auswirkungen Desinformationen auf

die Einstellung der Rezipienten haben. Seit Falschinformationen allerdings nicht mehr nur ein Phänomen sind, das gelegentlich (zum Beispiel aufgrund ungenauer Recherchen) in massenmedialen Kontexten auftaucht, sondern Social Networks betrifft, werden vermehrt auch Auswirkungen auf Verhalten untersucht – im Sinne der Frage, inwieweit Rezipientinnen durch ihr Weiterleitungsverhalten für die Verbreitung von Desinformation sorgen. Seitdem liegt der Fokus der Forschung auf Medien wie Facebook und Twitter. Da die Bekämpfung der Verbreitung von und Einflussnahme durch Desinformation ein Anliegen sowohl der Bevölkerung als auch von Expertenseite ist, wurden verschiedene Interventionsmöglichkeiten und deren Wirksamkeit untersucht. Im Folgenden wird ein Überblick über die bisherige Forschung aus medienpsychologischer Sicht und deren Ergebnisse mit Blick auf die genannten Aspekte gegeben.

I. Wirkung von Desinformation auf Einstellungen und Verhalten

1. Einfluss auf Einstellungen

Die Tatsache, dass jeder Online-Nutzende Inhalte produzieren und potentiell mit einer unbegrenzten Menge an anderen Nutzenden teilen kann, führt zu hohen Mengen an Informationen, mit denen Rezipientinnen konfrontiert werden. Im Regelfall kann als ein Ziel der Internetnutzung angenommen werden, dass Menschen relevante und korrekte Informationen suchen, um Entscheidungen daraus abzuleiten (zu Gesundheitsfragen wie ob bei gegebenen Symptomen ein Arzt aufgesucht werden sollte, zu Erziehungsfragen wie der potenziellen Gefahr durch zu starke Mediennutzung, zu Umweltfragen wie dem tatsächlichen Zustand des Weltklimas und den abzuleitenden Schlussfolgerungen, Co2 vermeidend zu leben, etc.). Ein Faktor bei der Auswahl der zu rezipierenden Nachrichten aus der Vielzahl an Informationen ist dabei, als wie glaubwürdig eine Nachricht empfunden wird.

Bereits in klassischen Studien zur Glaubwürdigkeit und Persuasion konnte identifiziert werden, dass die Quelle der Botschaft einen starken Einfluss auf Einstellungsbildung der Zuhörerschaft nimmt (Hovland, Janis & Kelley, 1953; Hovland & Weiss, 1951). Das liegt daran, dass die Quelle meistens nicht nur der salienteste Hinweisreiz ist, sondern Rezipienten darüber hinaus daran gewöhnt sind, die Quelle einer Information heranziehen, um deren Nützlichkeit einzuschätzen (Sundar, 2008). Diese Erkenntnisse zum Einfluss von Quellenglaubwürdigkeit konnten auch auf den Social-Network-Kontext

transferiert werden, da sich gezeigt hat, dass Nutzende beobachtbare Aspekte der Quelle heranziehen, um die Validität einer Botschaft einzuschätzen (Metzger, Flanagin & Medders, 2010). Dabei zeigt sich, dass Rezipienten durchaus kompetent vorgehen, indem sie die Nachrichten von Social-Network-Autoren (in diesem Fall Bloggerinnen) dann eher zum Lesen auswählen, wenn es Hinweise auf eine Expertise hinsichtlich des gegebenen Themenbereiches gibt (Winter & Krämer, 2012; vgl. auch Reputationsheuristik, Metzger et al., 2010). Dieser Effekt ist bei Personen, die ein hohes Bedürfnis haben, über Sachverhalte gründlich nachzudenken, besonders ausgeprägt.

Allerdings sind Rezipienten aufgrund von limitierten kognitiven Kapazitäten nicht jederzeit in der Lage, jede Information in einer elaborierten Form zu verarbeiten (Lang, 2000). Dies wird beispielsweise nachvollzogen im Rahmen verschiedener zwei-Wege-Modelle der Informationsverarbeitung und Einstellungsänderung. Im Elaboration Likelihood Modell (Petty & Cacioppo, 1986) und im Heuristic Systematic Modell (Chaiken, 1989) werden zwei unterschiedliche Wege beschrieben, über die Informationen rezipiert und weitergehend verarbeitet werden und dementsprechend Einfluss auf Einstellungen und Verhalten nehmen können. In Abhängigkeit von Fähigkeit und Motivation (und thematischem oder situationalem Involvement) wird entweder die zentrale Route, die eine elaborierte Evaluation und Verarbeitung aller eingehenden Informationen beinhaltet, oder die periphere Route gewählt, die eine eher oberflächliche Verarbeitung beschreibt, während der einfache Hinweisreize oder heuristische Regeln Eingang in die Einstellungsbildung finden können.

Im Zuge einer peripheren Verarbeitung erhöht sich die Wahrscheinlichkeit, dass manipulierte oder falsche Informationen, die durch andere Nutzende empfohlen oder beworben werden (beispielsweise auf Social-Network-Plattformen durch eine hohe Anzahl Likes und Shares, die als aggregierte Summe unter jedem Beitrag angezeigt wird), als glaubwürdig empfunden und daraus resultierend inhaltlich angenommen werden. Dieser Effekt begründet sich durch die implizite Annahme, dass „etwas, das von anderen als gut empfunden wird, von einem selbst auch als gut beurteilt werden kann". Diese Annahme wird als Bandwagon Heuristic (Sundar, 2008) beschrieben, die bereits für die Beurteilung von Online Reviews als einflussreich bestätigt werden konnte (Sundar, Oeldorf-Hirsch & Xu, 2008). Auch für das Vertrauen in inkorrekte Informationen kann dieser Effekt als möglicher Erklärungsansatz herangezogen werden.

Doch auch unabhängig von diesen Tendenzen, sich bei fehlender Fähigkeit oder fehlender Motivation mit den geposteten Nachrichten nicht tiefgehend

auseinanderzusetzen, lassen sich subtile kognitive Effekte nachweisen, denen der Mensch generell unterliegt. In diesem Sinne haben Pennycook, Cannon und Rand (2018) herausgefunden, dass die Tendenz, Desinformationen zu glauben, durch den Eindruck von Vertrautheit infolge einer vorherigen Rezeption gefördert wird. Durch ein gelungenes experimentalpsychologisches Design, in dem Rezipienten manche (Falsch)Nachrichten mehrfach und andere nur einmal präsentiert wurden, konnte nachgewiesen werden, dass falsche Nachrichten dann eher geglaubt werden, wenn man sie mehrfach sieht. Es ist anzunehmen, dass dieser Prozess der Fehlattribution unbewusst stattfindet, da auch Warnmeldungen keine Änderung in der Bewertung der Glaubwürdigkeit von bereits gesehenen Inhalten bewirkten. So wurden Fake-News-Überschriften, die zuvor bereits den Versuchspersonen präsentiert wurden, nach einer zweiten Rezeption als glaubwürdiger eingeschätzt, auch wenn sie von einer Warnung begleitet wurden. Fake-News-Überschriften, die zum ersten Mal und ohne Warnmeldung gezeigt wurden, wurden hingegen als weniger glaubwürdig bewertet.

Ein weiterer kognitiver Fehler, dem alle Nutzenden unterliegen, bezieht sich darauf, dass Menschen nach Konsistenz in ihren Einstellungen, ihrem Verhalten und ihrer Selbstwahrnehmung streben und dadurch Informationen bevorzugen, die sich mit ihrer bereits bestehenden Meinung decken. Werden Nachrichtenkonsumierende vor die Wahl zwischen unterschiedlichen Artikeln und Quellen gestellt, dann entscheiden sie sich für diejenigen, die den eigenen Einstellungen und Ansichten am ehesten entsprechen und somit die geringste Wahrscheinlichkeit einer unangenehmen, Stress verursachenden kognitiven Dissonanz beinhalten (vgl. cognitive dissonance theory, Festinger, 1962). Dieses Phänomen wird auch als Confirmation Bias bezeichnet (Nickerson, 1998). Neben der Tatsache, dass diese Tendenz als eine mögliche Ursache für das Zustandekommen sogenannter Filterblasen diskutiert wird (Lazer et al., 2017), wird auch argumentiert, dass der Confirmation Bias die Wirkung von Falschinformationen verstärken kann – wenn Nutzende eine Voreinstallung haben, die dazu führt, dass sie die Falschinformationen glauben wollen. Der Confirmation Bias wird sogar wirksam, wenn Desinformation als falsch enttarnt wird (Tan & Ang, 2017; Ott, 2017).

2. Einfluss auf Verhalten – Weiterleitung von Desinformation

In Bezug auf Desinformationen in Social Networks ist auch ein spezifischer Einfluss auf Verhalten äußerst relevant. Desinformationen verbreiten sich

nämlich dann besonders rasant, wenn sie von Menschen weitergeleitet werden. Durch die Merkmale von Social Networks wie hohe Vernetzung und Einfachheit sowie Schnelligkeit der Kommunikation (bedingt durch einfache Mechanismen wie liken und sharen: Aktionen, die mit einem Klick möglich sind und potenziell eine große Reichweite haben können; vgl. das Konzept der Mass Interpersonal Persuasion von Fogg, 2008) wird begründet, dass auch falsche Nachrichten sich durch Tausende von Likes schnell viral verbreiten. Das wiederum kann im Weiteren eine irreführende Illusion von Vertrauen in bestimmte, vielfach geteilte (und damit, in Anlehnung an gängige Social-Network-Kommunikationskonventionen, empfohlene) Inhalte erzeugen (Elyashar, Bendahan & Puzis, 2017). In diesem Zusammenhang zeigt eine Studie von Vosoughi, Roy und Aral (2018), dass sich Falschmeldungen in Social Networks sogar schneller und häufiger verbreiten als wahre Meldungen, besonders bei politischen Themen. Die Top 1 % der beliebtesten Falschnachrichten erreichten demnach 1.000 bis 100.000 Menschen, während wahre Meldungen selten mehr als 1.000 Menschen erreichten. Die zentrale Erkenntnis der Studie besteht allerdings darin, dass diese Verbreitung eher durch Menschen als durch Bots entsteht. Dabei stellen die Autoren die Hypothese auf, dass insbesondere die Neuartigkeit der Nachrichten ein relevanter Faktor für die Weiterleitung darstellt. Dies sollte allerdings in weiteren Studien geprüft werden, denn weitere Einflussfaktoren könnten etwa der Negativity Bias (die Tatsache, dass negative Nachrichten mehr Aufmerksamkeit auf sich ziehen) oder die Emotionalität der Nachrichten sein. Oftmals werden Falschnachrichten so konzipiert und kommuniziert, dass sie die Emotionen und Gefühle der Rezipientinnen ansprechen anstatt Fakten zu transportieren. Während der US-Wahlen 2016 hat sich beispielsweise gezeigt, dass emotional dargestellte Geschichten gegenüber Fakten bevorzugt wurden (Gross, 2017). Auf der anderen Seite muss gefragt werden, wie viel Einfluss die Inhalte der Nachrichten überhaupt haben. Andere Studien finden nämlich, dass 59 % der Links auf Twitter geteilt werden, ohne aufgerufen worden zu sein (Gabielkov et al., 2016).

Der Einfluss von Falschnachrichten kann sich gegebenenfalls dadurch verstärken, dass Social-Network-Kanäle oftmals als einzige Quelle für den Konsum von News und politischen Inhalten herangezogen werden, ohne dass weitere Quellen wie Zeitungen und Magazine konsultiert werden. Daraus resultierend sind Rezipienten heutzutage (bzw. im Social-Network-Kontext) weitaus anfälliger für Manipulation und falsche Informationen.

II. Einfluss von Interventionen

Wie eine aktuelle Bevölkerungsumfrage zeigt, wünschen sich viele Deutsche, dass Desinformationen besser und flächendeckender bekämpft werden (PricewaterhouseCoopers GmbH, 2019). Obwohl die meisten Social-Networks-Nutzende zuerst einmal versuchen, Desinformation selbstständig zu erkennen und sich so vor einem Konsum und entsprechender Beeinflussung zu schützen, wünschen sie sich darüber hinaus Schutz durch externe Quellen zur Erkennung von Falschinformation (vgl. Tandoc et al., 2017). Als geeignete Gegenmaßnahmen werden dabei vor allem eine Löschung von Inhalten durch Netzwerkbetreiber und die Aufklärung der Bevölkerung durch die Regierung genannt (PricewaterhouseCoopers GmbH, 2019). Des Weiteren sehen Bürgerinnen die Medien in der Pflicht, Falschinformationen zu identifizieren und zu korrigieren. Tatsächlich wird seit einigen Jahren verstärkt auch psychologische Forschung zu geeigneten Gegenmaßnahmen und deren Wirksamkeit durchgeführt. In den folgenden Abschnitten wird eine Einordnung von Interventionsmöglichkeiten unter psychologischen Gesichtspunkten vorgenommen sowie jeweils eine Auswahl an konkreten Maßnahmen mit empirischen Ergebnissen zu deren Wirkung gegeben.

1. Nachhaltigkeit der Auswirkungen von Desinformation: Falschinformationseffekt

Bereits vor dem Aufkommen von Social Networks existierten Desinformationen, die man mit Hilfe von Korrekturen zu bekämpfen versuchte. In diesem Zusammenhang wurde allerdings schon in den 1970er Jahren in der Psychologie der sogenannte Falschinformationseffekt beschrieben, in dessen Rahmen erkannt wurde, dass sich einmal rezipierte Information nicht einfach korrigieren lässt. Zahlreiche Studien belegen, dass auch nach Aufklärung über die falsche Information und auch unter Bedingungen, in denen Probanden sich eigentlich darüber im Klaren sind, dass die ursprünglich gegebene Information falsch oder wenig vertrauenswürdig ist (Ross, Lepper & Hubbard, 1975) und sie motiviert sind, die neue Information zu glauben, die alte, falsche Information im Gedächtnis bleibt und abgerufen wird. Es handelt sich daher nicht um ein Problem unzureichender Beachtung der neuen Information oder ein motivationales Problem (zum Beispiel der neuen korrekten Information keinen Glauben schenken zu wollen), sondern um einen rein kognitiv erklärbaren Effekt, der auf die Funktionsweise des Gehirns zurückzufüh-

ren ist. Die ursprüngliche, falsche Information wird nämlich in bestehende Wissensbestände plausibel eingebaut und mit dieser so vernetzt, dass ein Überschreiben durch die neue Information nur in Ausnahmefällen gelingt (Graesser, Singer & Trabasso, 1994). Insbesondere, wenn Erklärungen für die Sachverhalte gefunden werden, wird an diesen auch noch festgehalten, wenn über die Falschheit informiert wurde (Anderson, Lepper & Ross, 1980). Nur wenn die neue, korrigierte Information die alte Information im gebildeten Modell problemlos ersetzen kann, tritt kein Falschinformationseffekt auf (Johnson & Seifert, 1994). Ein essentielles Ergebnis in der Forschung zum Thema Falschinformationseffekt ist somit, dass die Identifizierung und Korrektur von Falschnachrichten aufgrund von subtilen kognitiven Effekten nicht zwangsläufig zu einer Meinungsänderung in Richtung der korrekten Information führen, da die Veränderung eines bereits gebildeten und mitunter durch ähnliche Informationen bestätigten Eindrucks schwierig ist. Gegenmaßnahmen sind zusätzlich insbesondere im Bereich von Social Networks durch das dynamische Umfeld erschwert, denn die Informationen erscheinen nicht nur an einer Stelle und können gegebenenfalls nicht überall mit der Korrektur versehen werden. Denn sobald eine Information online veröffentlicht ist, kann angenommen werden, dass „sie bereits Beine hatte" ("the story already had legs", Berghel, 2017).

2. Gegenmaßnahmen

Eine Aufklärung der Gesellschaft über Desinformation stellt den ersten Schritt zur Bekämpfung von Falschinformationsverbreitung dar. So konnte bereits gezeigt werden, dass eine allgemeine Warnung vor Desinformation hilft, um die Verbreitung von Falschinformationen zu mindern (Clayton et al., 2019). Insbesondere ältere Menschen scheinen Bedarf für solche Maßnahmen zu haben, da gezeigt wurde, dass sie auf Social-Network-Plattformen häufiger Falschinformationen teilen als jüngere Bevölkerungsgruppen (Guess, Nagler & Tucker, 2019). Da sich außerdem gezeigt hat, dass die Fähigkeit zum analytischen und strukturierten Denken das Erkennen von Falschinformationen begünstigt (Pennycook & Rand, 2018), sollten Maßnahmen getroffen werden, um diese Art der Informationsverarbeitung zu fördern.

Allerdings ist das Erkennen von Falschinformationen für Laien selbst mit Hintergrundwissen und trotz einer generellen Sensibilisierung schwierig. Daher müssen geeignete Gegenmaßnahmen eingesetzt werden, um Bürgerinnen vor dem Einfluss von Falschinformation zu schützen. Diese Interventionen

bestehen in drei wesentlichen Maßnahmen, die getrennt, aber auch in Kombination angewendet werden können. Zum einen gibt es die Möglichkeit, Desinformationen zu löschen, um sie von einem weiteren Konsum beziehungsweise einer umfassenderen Verbreitung auszuschließen. Zum anderen können Warnhinweise installiert werden, die Inhalte als desinformierend oder zumindest teilweise fehlleitend kennzeichnen. Außerdem kann nach der erfolgreichen Erkennung von Falschinformationen eine Korrektur in Form einer Gegendarstellung vorgenommen werden.

Löschung von Desinformation

Die erste bekannte Maßnahme, um Fake News aktiv entgegenzuwirken, besteht in der Löschung solcher Meldungen. Natürlich müssen dafür desinformierende Inhalte zuerst einmal erkannt werden. Dann stehen Betreiber von Netzwerkseiten und anderen Social-Network-Portalen in der Pflicht, diese Inhalte zu löschen. Da bei der Menge an Informationen im Netz eine flächendeckende manuelle Sichtung ausgeschlossen ist, werden automatisierte Verfahren erarbeitet (vgl. Kapitel 4). Da eine zweifelsfreie Erkennung von anderen Inhalten wie beispielsweise Satire oder Ironie nicht einfach ist, ist allerdings eine direkte Löschung häufig problematisch.

Zudem kann eine Löschung immer erst dann passieren, wenn ein Beitrag bereits veröffentlicht wurde. Besser wäre es, wenn effizientere und effektivere Maßnahmen verfügbar wären, die die Verbreitung von Falschnachrichten von Vornherein verhindern. Wenn Informationen erst einmal im Internet geteilt werden, werden sie zwangsläufig von vielen Menschen konsumiert. Eine anschließende Löschung kann also nicht verhindern, dass Desinformation trotzdem schon von vielen Rezipientinnen gelesen und verinnerlicht wurden.

Darüber hinaus erscheint das Bekämpfen von Falschmeldungen am wirksamsten zu sein, wenn eine eindrückliche Gegendarstellung mit wahren Argumenten und Fakten geboten wird (s.u.). Fehlt eine solche, wird die ursprüngliche, fehlerhafte Information nicht überschrieben und bleibt trotzdem im Wissensbestand erhalten. Löschungen sind also nur dann sinnvoll, wenn diese unmittelbar nach der Veröffentlichung eines Postings geschehen. So kann die Anzahl an Rezipienten eingeschränkt und ein Weiterleiten von Falschnachrichten verhindert werden. Besser ist es, wenn Falschmeldungen gar nicht erst veröffentlicht werden oder auch Quellen, die für das Verbreiten von Falschinformationen bekannt sind, blockiert werden (vgl. Lazar et al., 2017).

Warnhinweise für Falschinformationen

Warnhinweise sind allgegenwärtig und finden Anwendung in verschiedensten Kontexten sowie Disziplinen. So finden sie sich im Straßenverkehr, am Arbeitsplatz oder auf Verpackungen mit gesundheitsschädlichem Inhalt. Sie dienen dazu, die allgemeine Sicherheit zu unterstützen, Informationen über mögliche Gefahren bereitzustellen, das Verhalten von Personen zu beeinflussen oder eine Gefahr ins Gedächtnis zurückzurufen (Laughery & Wogalter, 2006). Eine Voraussetzung für die erfolgreiche Wirkung von Hinweisen ist zuerst einmal, dass sie als solche erkannt werden. Deswegen sollten bestimmte Gestaltungsrichtlinien, die ihren Ursprung in der Wahrnehmungspsychologie haben, eingehalten werden.

Damit Warnungen wirken können, muss deren semantische Bedeutung für alle Menschen klar erkenntlich sein. So ist es in der Regel hilfreich, nicht ausschließlich auf den Einsatz von Symbolen oder Text zu vertrauen, sondern eine Kombination von beidem anzustreben (Haynes, 2016). Ein eindeutiges Icon für die Identifikation von Desinformation könnte also mit einer entsprechenden Erklärung beziehungsweise mit einem einzigen Signalwort (z. B. Gefahr oder Warnung) kombiniert werden (vgl. Laughery & Wogalter, 2006). Außerdem sind eindrückliche Farben (z. B. signalrot) eine gute Ergänzung, da sie Aufmerksamkeit auf sich ziehen und sich von der Umgebung absetzen (Laughery & Wogalter, 2006). Weiterhin ist es wichtig, ein Warnsymbol stets in der unmittelbaren Nähe der Gefahr anzubringen (Wogalter, Conzola & Smith-Jackson, 2002). Noch besser wirken Warnungen, wenn damit interagiert werden muss, sie also aktiv beseitigt werden müssen, um ein Produkt o.ä. zu nutzen (Laughery & Wogalter, 2006). Im Kontext von Digitaler Desinformation könnte das ein Pop-Up Fenster mit einer Warnung vor Falschinformationen sein, das weggeklickt werden muss.

Einige Studien haben sich bereits mit spezifischen Warnhinweisen beschäftigt und diese evaluiert. Die Ergebnisse zu deren Wirkungen sind allerdings teilweise widersprüchlich. So scheinen laut Geary (2017) und Gao, Xiao, Karahalios und Fu (2018) sogenannte Credibility Cues auf Facebook nicht wirksam zu sein. Diese Cues wurden direkt an Desinformations-Postings angehängt, um diese als Falschinformation zu kennzeichnen. Dieses Vorgehen entspricht also den empfohlenen Gestaltungsempfehlungen für Warnhinweise, da Gefahr und Warnung direkt nebeneinander sichtbar waren. Allerdings wurden sie dennoch nicht richtig wahrgenommen, da Nutzende typischerweise schnell durch den Facebook-News-Feed scrollen und so nur wenige der verfügbaren Informationen verarbeiten. Des Weiteren kann das Kennzeich-

nen von Beiträgen dann problematisch sein, wenn nur ein Teil der Postings als falsch gekennzeichnet wird. Dies impliziert, dass der Inhalt der übrigen, nicht gekennzeichneten Beiträge wahr ist, obwohl auch diese falsche Informationen enthalten können (implied truth effect, Pennycook et al., 2019).

Andererseits konnte gezeigt werden, dass Markierungen, die falschen („rated false") oder zumindest umstrittenen („disputed") Inhalt anzeigen, wirksam gegen Desinformation sind. Sie sorgen dafür, dass Überschriften desinformierender Artikel als weniger akkurat eingeschätzt werden. Die Wirksamkeit dieser Markierungen ist sogar dann vorhanden, wenn die im kritisierten Artikel ausgedrückte Meinung der der Probandinnen entsprach (Clayton et al., 2019). Außerdem sind sogenannte Truth Scales („Ampeln", die den Wahrheitsgehalt einer Nachricht anzeigen) laut Amazeen, Thorson, Muddiman und Graves (2018) wirkungsvoll, um Fake News als falsch zu kennzeichnen. Darüber hinaus sorgen sie sogar nachhaltig dafür, dass Falschnachrichten langfristig als solche abgespeichert werden. Allerdings müssen Truth Scales zuerst einmal von Lesern gefunden werden, da sie meist auf speziellen Fact-Checking Webseiten zur Verfügung stehen. Nutzende müssen also aktiv werden, um sich über Falschinformationen zu informieren.

Abschließend ist es wichtig zu bemerken, dass Warnhinweise generell sehr bedacht gestaltet werden sollten. Verschiedene Studien haben gezeigt, dass eine fehlerhafte Gestaltung solcher Hinweise sogar dazu führen kann, dass Falschnachrichten als wahre Nachrichten abgespeichert werden (Skurnik et al., 2005; Wogalter et al., 2002). Sie erregen dann zwar Aufmerksamkeit, korrigieren den Fehleindruck aber nicht ausreichend und verankern den Inhalt der irreführenden Nachricht sogar stärker im Gedächtnis. Deswegen sollte die Wirksamkeit spezifischer Warnhinweise stets empirisch getestet werden.

Korrekturen und Gegendarstellungen von Falschinformationen

Die letzte und wirksamste Möglichkeit, Fake News zu bekämpfen, besteht darin, diese an geeigneter Stelle zu korrigieren. Das Verhindern einer Verbreitung von Falschinformationen ist zwar das beste Mittel, um Falschnachrichten zu begegnen, doch ist dies nicht immer möglich. Korrekturen können hingegen so gestaltet werden, dass sie eine hohe Chance auf Erfolg haben. Im Folgenden werden verschiedene Möglichkeiten für eine effektive Gestaltung sowie diverse Probleme, die sich bei der Verbreitung von Korrekturen ergeben können, beschrieben.

Bisher konnte gezeigt werden, dass das Eindämmen einer Falschinformationsverbreitung wirksamer ist als jede nachträgliche Korrektur. So gelangen Korrekturen beispielsweise längst nicht an alle Leser der ursprünglichen Falschnachricht. Darüber hinaus verbreiten sich Korrekturen generell deutlich weniger als Desinformationen.

Da hauptsächlich Menschen - und nicht etwa Social Bots - für die Verbreitung von Falschnachrichten verantwortlich sind (Vosoughi et al., 2018), sollte eine Weiterleitung von Falschnachrichten durch Social-Networks-Nutzende bekämpft werden. Eine Korrektur des Beitrags ist nur dann sinnvoll, wenn diese den ursprünglichen Link zum desinformierenden Nachrichtenartikel nicht enthält. Algorithmen auf sozialen Netzwerken erkennen, wenn ein Link weiterhin geteilt wird, und empfehlen entsprechende Inhalte an andere Nutzende weiter. Ein viel gelesener Beitrag taucht so beispielsweise vermehrt im Facebook-News-Feed auf. In ähnlicher Weise ist es auch nicht empfehlenswert, eine Information als fehlleitend zu markieren und auf dem eigenen Social-Network-Profil zu teilen, da sie von anderen Social-Network-Nutzenden trotzdem als wahr abgespeichert werden könnte. Social-Network-Nutzende nehmen Informationen häufig nur am Rande wahr und könnten so einen Hinweis auf Falschnachrichten übersehen. Besonders hilfreich scheint es hingegen zu sein, wenn Freunde in Social Networks darauf hinweisen, wenn jemand Falschnachrichten (auch ohne bewusste Intention) geteilt hat. Die Akzeptanz einer solchen Korrektur ist höher als die einer unbekannten Quelle. Außerdem werden Richtigstellungen durch Quellen, die mit der eigenen politischen Meinung übereinstimmen, eher angenommen und verinnerlicht (Lazer et al., 2017). Bemerkenswert ist trotzdem, dass es sich generell schwierig gestaltet, eine Falschinformation zu überschreiben, vor allem, wenn diese mit vorherigen Einstellungen einer Person übereinstimmt (Nyhan & Reifler, 2010; vgl. Falschinformationseffekt, Kapitel 3, B.I.).

Daher ist eine Aufgabe für Fact-Checking Organisationen oder Journalistinnen, ansprechende Narrative für ihre Korrekturen zu finden. Diese können dazu beitragen, dass sich korrigierende Informationen schnell und weit verbreiten, da sie das Interesse und die Aufmerksamkeit von Lesern wecken. Darüber hinaus kann eine involvierende Geschichte, in der korrekte Alternativerklärungen und Hintergrundinformationen geboten werden, dafür sorgen, dass sich richtige Information fester in den vorhandenen Wissensstand einfügen.

Bei der Widerlegung von Fake News sollten generelle Regeln des sogenannten „Debunking“ eingesetzt werden, um diese wirksam zu gestalten (Cook & Lewandowsky, 2011). Das Berichten von Fakten sollte bei einer

Korrekturmeldung generell im Vordergrund stehen, komplizierte Formulierungen sollten vermieden werden und Alternativerklärungen für berichtete Ereignisse sollten enthalten sein. Nyhan und Reifler (2015) konnten zeigen, dass es nicht ausreicht, eine Desinformation über das Auftreten eines Ereignisses lediglich zu negieren. Leserinnen benötigen eine kausale Begründung, um die korrekte Information abzuspeichern. Außerdem sollten unnötige Wiederholungen von Falschnachrichten vermieden werden, um sie nicht im Wissensnetz zu festigen. Allerdings wurde gezeigt, dass durch eine explizite Bezugnahme und Wiederholung der ursprünglichen Falschinformation innerhalb einer korrigierenden Gegendarstellung das Vertrauen in die Falschinformation signifikant stärker gesenkt werden konnte als durch eine Korrekturdarstellung ohne Wiederholung (Ecker, Hogan & Lewandowsky, 2017). Das heißt, dass solange beim Verfassen einer Gegendarstellung der Fokus auf der Bestätigung von Tatsachen statt auf der Widerlegung von Mythen liegt, eine Wiederholung der ursprünglichen Falschinformation ihre Wirkung sogar zu reduzieren vermag (Ecker, Hogan & Lewandowsky, 2017).

Teilweise kann eine Korrektur auch dann besonders wirksam sein, wenn sie in Verbindung mit Warnhinweisen oder in einer speziellen Darstellungsform veröffentlicht wird. Grafische Wahrheitsskalen (Truth Scales, s.o.) können korrigierende Informationen unterstützen und verfestigen (Amazeen et al., 2018). Auch erklärende Texte sind in der Lage Fehlinformationen erfolgreich zu widerlegen (Amazeen et al., 2018). Des Weiteren können Videos als Medium eingesetzt werden, um die Aufmerksamkeit für korrekte Informationen zu steigern und eine mögliche Verwirrung über Ereignisse zu verringern (Young, Jamieson, Poulsen & Goldring, 2017).

C. Zusammenfassung eigener Forschung

Aufbauend auf den oben zusammengefassten Erkenntnissen haben wir durch eigene Studien zu einer Verbreiterung der Wissensbasis beigetragen. Zuerst wurde untersucht unter welchen Bedingungen Falschinformationen einen Einfluss haben können. Hierzu wurden Merkmale der Nachricht, der Quelle und der Rezipientinnen betrachtet. Außerdem wurde erforscht, wie Interventionen gestaltet werden sollten, um eine möglichst wirksame Bekämpfung von Falschinformationen zu erreichen.

I. Unter welchen Bedingungen ist Desinformation einflussreich?

Bevor die Gestaltung verschiedener Interventionsmöglichkeiten und deren Wirkungsweise beleuchtet werden, wurden Grundlagen für die Implementierung solcher Gegenmaßnahmen untersucht. Dazu wurden vor allem spezifische Eigenschaften von Nachrichten und deren Einfluss auf Einstellungen untersucht. Ein weiterer Fokus lag auf Charakteristika der Quellen, die Falschnachrichten verbreiten, sowie auf den Persönlichkeitsmerkmalen von Rezipienten.

1. Einfluss von Nachrichtenmerkmalen auf resultierende Einstellungen

Um zu adressieren und zu überprüfen, inwieweit verschiedene Nachrichtenmerkmale von Falschinformationen die wahrgenommene Glaubwürdigkeit von fabrizierten Online-Nachrichten und die Weiterverbreitungsintentionen beeinflussen, wurde eine experimentelle Studie konzipiert und durchgeführt. In dieser wurde den Studienteilnehmenden ein Online-Nachrichtenartikel präsentiert, der eines der folgenden fünf Merkmale von Desinformation aufwies: reißerische Formulierungen, Inkonsistenzen, Subjektivität, unglaubwürdige Quelle oder manipuliertes Bild. In einer Kontrollbedingung wurde der Basis-Text ohne eine dieser Merkmalsvariationen präsentiert, sodass es insgesamt sechs verschiedene Versuchsbedingungen gab. Um Erkenntnisse zu verschiedenen Themenbereichen zu gewinnen, wurden zwei verschiedene Nachrichtenartikel (Kriminalität und Pflege) verwendet. Gemessen wurde, als wie glaubwürdig und akkurat die Teilnehmenden den Artikel einschätzten und mit welcher Wahrscheinlichkeit sie ihn weiterleiten würden. Insgesamt nahmen 294 Personen an der Studie teil. Die Ergebnisse des Experiments zeigen, dass keines der untersuchten Nachrichtenmerkmale (z. B. reißerische Formulierungen) in besonderem Maße dazu führt, dass Desinformation weniger glaubwürdig bewertet wird oder eher weitergeleitet wird. Generell begegneten die Probandinnen den Nachrichten aber mir eher großer Skepsis.

In zwei weiteren experimentellen Studien wurde die Valenz der Nachricht (positiv / negativ) auf die Wirkung einer politischen Desinformation untersucht. Dafür wurde eine positive oder eine negative Falschnachricht über einen deutschen Politiker gezeigt. Dazu wurden verschiedene positive (z. B. Knochenmarkspende, Ausrichtung einer Benefizgala) und negative (z. B. Steuerhinterziehung, Schlägerei) Falschnachrichten generiert, vorgetestet und randomisiert im Experiment präsentiert. Gemessen wurde die Bewer-

tung des Politikers, vor und nachdem die Probanden die Information erhielten, dass die Nachricht erfunden war. In der ersten Studie wurde diese Korrektur durch die Versuchsleiterin in Form einer mündlichen Aufklärung vorgenommen. Im zweiten Szenario bekamen Probandinnen eine Richtigstellung als Meldung einer Fact-Checking Organisation zu sehen.

Die Ergebnisse zeigen, dass positive Falschinformationen glaubwürdiger bewertet werden als negative, negative allerdings brisanter wahrgenommen werden. Die Bewertung des Politikers wird marginal durch die Valenz der Nachricht beeinflusst, sodass eine negative Nachricht zu einer schlechteren Bewertung führt. Es gibt somit Hinweise darauf, dass negative Falschmeldungen folgenschwerer sein könnten, dies muss aber in weiteren Studien geprüft werden.

2. Einfluss von Attributen der Quelle

Des Weiteren wurde der Einfluss der Informationsquelle einer in sozialen Medien geteilten Nachricht untersucht. Der genaue Fokus lag hierbei auf dem Einfluss dem Vermittler einer Botschaft mit besonderer Berücksichtigung seiner Beziehung zur Rezipientin. Um die spezifische Umgebung von Social-Networks-Websites bei der Frage nach der Wirkung online geteilter Falschinformationen zu berücksichtigen, wurde untersucht, ob es einen Unterschied macht, wenn eine Falschinformation von einer guten Freundin oder einem entfernten Bekannten auf Facebook geteilt wurde. Dazu wurden den Studienteilnehmenden individualisierte Nachrichten-Posts gezeigt, die augenscheinlich von einem ihrer engen oder losen Facebook-Freundinnen geteilt wurden. Dabei zeigte sich, dass die Beziehungsstärke zum teilenden Freund keinen Einfluss auf die Bewertung der Nachricht oder die Bewertung des Politikers hatte.

3. Einfluss von Personenmerkmalen

Um den Einfluss der Merkmale des Empfängers einer Botschaft zu untersuchen, wurden verschiedene Personenvariablen, wie z. B. die Tendenz zu analytischem Denken (Need for Cognition) und die Einstellung zum Thema der Nachricht, im Rahmen einer experimentellen Studie erfasst. Weiterhin wurde untersucht, als wie glaubwürdig und akkurat verschiedene Nachrich-

tenartikel von Studienteilnehmenden in Abhängigkeit ihrer Ausprägung der Persönlichkeitsvariablen bewertet wurden.

Individuelle Prädispositionen scheinen entscheidend für die Glaubwürdigkeitswahrnehmung sowie für die Einschätzung der Akkuratheit von Desinformation zu sein. Die Ergebnisse deuten darauf hin, dass insbesondere Personen mit einer ausgeprägten Neigung zum analytischen Denken (Need for Cognition) Desinformation als weniger akkurat einstufen und dass die eigene Meinung zum Thema der Nachricht ein starker Prädiktor für die wahrgenommene Glaubwürdigkeit von Falschnachrichten ist.

II. Wirkung von Gegenmaßnahmen

Nachdem Einflussfaktoren für die Wirkung von Falschinformationen evaluiert und analysiert wurden, wurden konkrete Interventionsmechanismen entwickelt. Dazu wurde zunächst untersucht, wie die Gestaltung von Warnhinweisen aussehen sollte. Anschließend wurde die Wirksamkeit von Falschmeldungskorrekturen ergründet sowie der Einfluss von Reaktionen anderer Nutzender erforscht.

1. Studien zu Warnhinweisen

Bezüglich der Wirkung von Warnhinweisen wurde überprüft, wie Desinformation verarbeitet wird, wenn sie als fehlleitend markiert wird. Um sich dieser Frage zu nähern, wurde zunächst eine Voruntersuchung zur Wahrnehmung von Warnhinweisen durchgeführt. Dabei wurden den Teilnehmenden in einem Online-Experiment verschiedene Arten von Warnhinweisen präsentiert und anschließend erhoben, wie die Nützlichkeit wahrgenommen wird. Variiert wurde einerseits die Grafik des Hinweises (Warndreieck vs. Prozentuale Skala) und andererseits die textuelle Formulierung („als falsch bewertet“ vs. „umstritten“, vgl. Clayton et al., 2019). Die Ergebnisse der Studie zeigen, dass ein „einfaches“ Warndreieck nützlicher und glaubwürdiger bewertet wurde als eine prozentuale Warnskala und zu einer höheren Akzeptanz führte. Die textuelle Variation hatte keinen Effekt auf die wahrgenommene Nützlichkeit oder Glaubwürdigkeit der Warnung. Als positive Effekte von Warnhinweisen merkten die Teilnehmenden an, dass sie zu einer intensiveren Beschäftigung mit dem Thema der Meldung führen würden und bestimmte Personengruppen, wie Kinder und Jugendliche, schützen können.

Negativ hervorgehoben wurde, dass es keine Quellenangabe zum Warnhinweis gab, sodass der Ursprung der Bewertung unklar blieb und das Vertrauen dadurch möglicherweise negativ beeinträchtigt wurde. Ein möglicher Grund für die schlechtere Bewertung der prozentualen Warnskala könnte sein, dass diese bei den Lesenden Unklarheit darüber auslöst, ob es einen richtigen und einen falschen Teil der Meldung gibt und welcher Anteil der Meldungen falsch und welcher richtig ist.

Aufbauend auf diesen ersten Erkenntnissen wurde ein Laborexperiment konzipiert, in dem mittels Videoaufzeichnung des Klickverhaltens von Teilnehmenden untersucht werden sollte, welche Wirkung Warnhinweise auf das Selektions- und Leseverhalten von Online-Nachrichtenartikeln haben. Dazu wurde eine klickbare Version einer Nachrichtenplattform erstellt, auf der den Teilnehmenden eine Auswahl verschiedener Nachrichtenüberschriften präsentiert wurde, von denen die Hälfte mit einem Warnhinweis (rotes Warndreieck) markiert war. Die Teilnehmenden wurden instruiert, innerhalb von vier Minuten die Nachrichten auf der Seite zum genaueren Lesen auszuwählen, die sie sich näher ansehen wollten. Um zu überprüfen, ob als falsch markierte Nachrichten besonders häufig ausgewählt werden, wenn der Warnhinweis salient ist, wurde zudem variiert, ob entweder nur die Hälfte der Überschriften mit einem Hinweis (roter Warnhinweis, als falsch markiert) versehen wurden oder ob auch die restlichen Überschriften einen Hinweis (grüner Hinweis, als wahr markiert) enthielten. Außerdem sollte untersucht werden, wie Warnhinweise in Kombination mit Hinweisen auf Bewertungen von anderen Nutzenden (Anzahl von Likes) wahrgenommen werden. Daher wurde zusätzlich variiert, ob die Überschriften eine hohe (ca. 9000) oder niedrige (ca. 100) Anzahl von Likes aufwiesen. Die Analyse der Daten zeigt, dass Artikel mit Warnhinweisen generell seltener ausgewählt wurden als Artikel ohne Warnhinweis. Wenn alle Artikel markiert waren (grüne und rote Hinweise), wurden häufiger als falsch markierte Nachrichten ausgewählt als wenn ausschließlich rote Warnhinweise verfügbar waren. Besonders saliente Hinweise, die fehlleitende Artikel gegenüber wahren Nachrichten speziell abgrenzen, scheinen also abschreckender zu sein, als wenn alle Artikel gekennzeichnet sind. Die Anzahl an Likes anderer Nutzender hatte wiederum keinen systematischen Einfluss auf die Häufigkeit, mit der bestimmte Artikel zum Lesen selektiert wurden.

2. Wirkung von Korrekturen

Um zu untersuchen wie Korrekturen (inhaltlich) gestaltet und Online-Nutzenden präsentiert werden sollen, um die Wirkung von Falschinformation zu reduzieren, wurde aufbauend auf den bisherigen Kenntnissen und systematischen Literaturrecherchen zu Wirkungen von Korrekturdarstellungen eine experimentelle Studie konzipiert. Um den Einfluss der Art der Korrektur zu untersuchen, wurde den Teilnehmenden eine Falschnachricht und entweder eine kurze, faktenbasierte Korrektur oder eine detaillierte Korrektur mit plausibler Begründung präsentiert. Außerdem wurde variiert, ob die Versuchsteilnehmenden beide Meldungen (die Falschnachricht und die Korrektur) gleichzeitig präsentiert bekommen oder die Korrektur einige Minuten verzögert gezeigt wird.

Die Ergebnisse dieser Studie zeigen, dass eine detailliertere Gegendarstellung nicht dafür sorgt, dass diese Korrektur häufiger geglaubt wird als die hierdurch widerlegte Falschnachricht. Allerdings konnte herausgefunden werden, dass die in der Korrektur enthaltenen Fakten zur wahren Begebenheit besser erinnert wurden und somit fester im Wissensnetz der Versuchspersonen verankert waren als wenn nur eine oberflächliche Korrektur verfügbar war. Zudem wurde gezeigt, dass eine detaillierte Korrektur eher dazu führt, dass wahre Fakten nachhaltig erinnert werden, wenn sie auf der gleichen Seite zu sehen ist wie die Falschnachricht, also zeitgleich konsumiert wird. Diese Resultate deuten darauf hin, dass es sinnvoll ist, eine Korrektur in Form eines plausiblen Gegennarrativs zur ursprünglich falschen Information zu gestalten. Darüber hinaus sollte sie unmittelbar in der Nähe der Desinformation dargestellt werden, damit Konsumierende von online Nachrichten wahre Details besser erinnern.

Auch die in Kapitel 3, B.I. aufgeführten Studien zur Wirkung von positiven und negativen Nachrichten lassen Rückschlüsse darauf zu, welche Wirkungen unterschiedliche Korrekturen bzw. unterschiedliche Korrekturquellen erzielen. Resultate beider Studien illustrieren, dass die Bewertung der Person nach der Aufklärung, dass es sich um eine falsche Information gehandelt hat, steigt, wenn die Nachricht negativ war. Bei positiven Nachrichten sind die Ergebnisse weniger eindeutig. Die erste Studie (in der die Information, dass es sich um eine erfundene Meldung handelte, durch die Versuchsleiterinnen gegeben wurde) zeigt, dass es keinen Effekt auf die Personenbewertung hat, wenn eine positive Nachricht korrigiert wird. Ergebnisse der zweiten Studie, in der die Korrektur scheinbar durch eine Fact-Checking-Organisation vorgenommen wurde, verdeutlicht hingegen, dass die Korrektur einer positiven Nachricht

dazu führen kann, dass die Person im Nachhinein negativer bewertet wird. In diesem Setting wird offensichtlich angenommen, dass der Politiker selbst für die Verbreitung der positiven Nachrichten gesorgt hat. Die Ergebnisse erweitern die Erkenntnisse zur Wahrnehmung und Wirkung von positiver und negativer Desinformation über Politikerinnen und zeigen insbesondere starke Effekte von Korrekturdarstellungen auf Social-Networks-Websites. Zusammengenommen deuten die Ergebnisse der beiden Experimente darauf hin, dass die Art und Weise, wie eine Korrektur präsentiert wird und von wem sie kommt, ein entscheidender Faktor dafür ist, dass der Einfluss von Desinformation reduziert werden kann.

3. Wirkung von Nutzerkommentaren und Ratings

Zusätzlich zu den Untersuchungen zur Wirkung von Warnhinweisen haben wir in einer Studie im Rahmen einer Masterarbeit untersucht, inwieweit Online-Nutzende selbst als Ressource zur Bekämpfung von Desinformation dienen können. Dazu wurden auf Basis eines Online-Experiments die Effekte von Nutzer-Kommentaren und numerischen Nutzer-Bewertungen der Artikelglaubwürdigkeit auf die Wahrnehmung von irreführenden Nachrichtenartikeln untersucht. Zudem wurde überprüft, inwieweit die Kommentare und Bewertungen anderer Nutzenden die Bereitschaft, den Artikel weiterzuleiten, beeinflussen. Die Ergebnisse zeigen, dass die Nutzer-Bewertungen und -Kommentare insgesamt wenig Einfluss auf die Glaubwürdigkeitswahrnehmung und Weiterleitungswahrscheinlichkeit von irreführenden Nachrichten haben. Negative Kommentare, die Zweifel gegenüber einem Online-Artikel ausdrücken, können jedoch die Glaubwürdigkeit von Falschnachrichten reduzieren, insbesondere für Personen, die sich selbst fähig fühlen, relevante Informationen in Social Networks zu finden. Außerdem zeigte sich, dass Personen weniger bereit dazu waren, den Artikel privat über Social-Messenger-Services weiterzuleiten, wenn negative Kommentare anderer Nutzende angezeigt wurden (Kluck et al., 2019).

D. Zusammenfassung

Insgesamt lässt sich vor dem Hintergrund des Forschungsstandes konstatieren, dass sich aus psychologischer Sicht einige Gefahren von Desinformation feststellen lassen, dass aber auch in verschiedener Hinsicht Grund zur Hoff-

nung besteht. So lässt sich nicht nur für einen vermutlich großen Teil der Bevölkerung feststellen, dass sie Desinformation vermeiden möchten, sondern auch, dass Studien zeigen, dass Personen größtenteils kompetent sind in der Auswahl von Informationen aus dem Internet. Dagegen stehen allerdings Befunde, dass Nachrichten oft auch ohne genauere Prüfung weitergeleitet werden und dass ein hoher Einfluss der Voreinstellung erkennbar ist. Auf Basis der Voreinstellung werden eben auch Nachrichten, die zwar den Verdacht erwecken, Falschinformationen zu sein, aber in das eigene Weltbild passen, rezipiert und gegebenenfalls verbreitet. Während es noch keine Lösungen für diese Auswirkungen des Confirmation Bias gibt, zeigen erste Studie zu Interventionen, dass es wirksame Möglichkeiten gibt, zumindest diejenigen, die keine gefestigte Meinung in Bezug auf die Falschinformation haben, davon abgehalten werden können, die Nachricht zu lesen, in die eigenen Einstellungen zu integrieren oder weiterzuleiten. Sowohl Warnmeldungen als auch Korrekturen können hilfreich sein und erste Studien geben Hinweise darauf, dass es wirksamere und weniger wirksame Interventionen gibt.

Literaturverzeichnis zu Kapitel 3

Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, 95(1), 28–48. https://doi.org/10.1177/1077699016678186

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of personality and social psychology*, 39(6), 1037.

Berghel, H. (2017). Lies, damn lies, and fake news. *Computer*, 50(2), 80–85. https://doi.org/10.1109/MC.2017.56

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., . . . Nyhan, B. (2019). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Political Behavior, 38(2), 173. https://doi.org/10.1007/s11109-019-09533-0

Cook, J., & Lewandowsky, S. (2012). The debunking handbook (Version 2). St. Lucia, Australia: University of Queensland.

Czerwinski, M., Lund, A., & Tan, D. (Eds.) 2008. CHI '08 Extended Abstracts on Human Factors in Computing Systems. New York, NY: ACM.

Ecker, U. K.H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? Journal of Applied Research in Memory and Cognition, 6(2), 185–192. https://doi.org/10.1016/j.jarmac.2017.01.014

Elyashar, A., Bendahan, J., & Puzis, R. (2017). Has the online discussion been manipulated? Quantifying online discussion authenticity within online social media. Retrieved from http://arxiv.org/pdf/1708.02763v2

Festinger, L. (1962). A theory of cognitive dissonance (Vol. 2). California, USA: Stanford University Press.

Gao, M., Xiao, Z., Karahalios, K., & Fu, W.-T. (2018). To label or not to label. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1–16. https://doi.org/10.1145/3274324

Geary, L. (2017). Spread of false news stories on Facebook: An assessment of credibility cues and personality (Master's Thesis), Morgantown, West Virginia.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological review, 101(3), 371.

Gross, M. (2017). The dangers of a post-truth world. Current Biology, 27(1), R1-R4. https://doi.org/10.1016/j.cub.2016.12.034

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science advances, 5(1), eaau4586.

Haynes, K. (2016). Investigating the methodology of warning symbol design (Master's Thesis), Auburn, Alabama.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion: Psychological studies of opinion change. Communication and persuasion: psychological studies of opinion change. New Haven, CT, US: Yale University Press.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. Public Opinion Quarterly, 15(4), 635. https://doi.org/10.1086/266350

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(6), 1420.

Kluck, J. P., Rösner, L., & Krämer, N. C. (in press) (2019) Doubters are more convincing than advocates - The impact of user comments and ratings on credibility perceptions of false news stories on social media. Studies in Communication and Media, 8(4).

Lang, A. (2000). The limited capacity model of mediated message processing. Journal of Communication, 50(1), 46–70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

Laughery, K. R., & Wogalter, M. S. (2006). Designing effective warnings. Reviews of Human Factors and Ergonomics, 2(1), 241–271. https://doi.org/10.1177/1557234X0600200109

Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). Combating fake news: An agenda for research and action.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. Journal of Communication, 60(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. Political Behavior, 32(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. Journal of Experimental Political Science, 2(1), 81–93. https://doi.org/10.1017/XPS.2014.22

Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. Critical Studies in Media Communication, 34(1), 59–68. https://doi.org/10.1080/15295036.2016.1266686

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. Journal of Experimental Psychology: General, 147(12), 1865-1880. https://doi.org/10.1037/xge0000465

Pennycook, G., & Rand, D. G. (2018). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking.

Pennycook, G., Bear, A., Collins, E., & Rand, D. G. (2019). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. http://dx.doi.org/10.2139/ssrn.3035384

PricewaterhouseCoopers GmbH. (2019). Fake News: Ergebnisse einer Bevölkerungsumfrage. Retrieved from https://www.pwc.de/de/technologie-medien-und-telekommunikation/PwC%20190420%20Berichtsband%20Fake%20News.pdf

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. Journal of personality and social psychology, 32(5), 880.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. Journal of Consumer Research, 31(4), 713–724. https://doi.org/10.1086/426605

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. Digital media, youth, and credibility, 73100.

Sundar, S. S., Oeldorf-Hirsch, A., & Xu, Q. (2008). The bandwagon effect of collaborative filtering technology. In M. Czerwinski, A. Lund, & D. Tan (Eds.), CHI '08 Extended Abstracts on Human Factors in Computing Systems (pp. 3453–3458). New York, NY: ACM. https://doi.org/10.1145/1358628.1358873

Tan, E. E. G., & Ang, B. (2017). Clickbait: Fake News and Role of the State. RSIS Commentaries, 026-17. Retrieved from https://dr.ntu.edu.sg/bitstream/10220/42108/1/CO17026.pdf

Tandoc, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2018). Audiences' acts of authentication in the age of fake news: A conceptual framework. New Media & Society, 20(8), 2745–2763. https://doi.org/10.1177/1461444817731756

Winter, S., & Krämer, N. C. (2012). Selecting science information in web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. Journal of Computer-Mediated Communication, 18(1), 80–96. https://doi.org/10.1111/j.1083-6101.2012.01596.x

Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. Applied Ergonomics, 33(3), 219–230. https://doi.org/10.1016/S0003-6870(02)00009-1

Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. Journalism & Mass Communication Quarterly, 95(1), 49–75. https://doi.org/10.1177/1077699017710453

Kapitel 4: Automatisierte Erkennung von Desinformationen

Autoren:

Oren Halvani
Wendy Freifrau Heereman von Zuydtwyck
Michael Herfert
Dr. Michael Kreutzer
Dr. Huajian Liu
Hervais-Clemence Simo Fhom
Prof. Dr. Martin Steinebach
Inna Vogel
Ruben Wolf
York Yannikos
Dr. Sascha Zmudzinski

Dieses Kapitel adressiert die Frage, wie Desinformationen automatisiert erkannt werden können. Die Notwendigkeit dieser Betrachtung leitet sich von der Menge an entsprechenden Inhalten ab, die potentiell erzeugt und verbreitet werden können. Ohne eine Möglichkeit, hier automatisch oder zumindest semi-automatisch Inhalte zu erkennen und zu filtern, sind die für entsprechende Vorgänge Verantwortlichen schnell überfordert.

Dabei wird von einem Anwender ausgegangen, der eine Beurteilung von eingehenden Meldungen durchführt und dabei technisch unterstützt wird. Dies adressiert insbesondere Redakteure und Medienschaffende. Es soll die Grundlage für ein Werkzeug oder treffender eine Sammlung von Werkzeugen geschaffen werden, mit denen Nutzende Hinweise für eine Manipulation der Inhalte einer Meldung oder ihre desinformierende Natur sammeln können. Diese Hinweise können beispielsweise darin bestehen, dass ein Foto bereits in einer früheren Pressemitteilung verwendet wurde und nun in einem anderen Kontext verwendet wird. Oder dass ein Foto Spuren von einer Bearbeitung aufweist, die möglicher Weise seine Aussage verändern. Das Erzeugen dieser Hinweise soll dabei automatisiert geschehen, die Interpretation hingegen muss heute noch durch Experten erfolgen. Die Technik assistiert also dem Anwender bei der Prüfung.

Die technischen Untersuchungen adressieren derzeit noch die Medientypen Text, Bild und Video jeweils für sich alleine. Jeder Medientyp weist dabei eigene Manipulationstypen und Erkennungsmethoden auf. Neben der Erkennung sind allerdings auch noch andere Aspekte zu beachten. Eine Automatisierung kann nur über eine Einbindung in ein Gesamtsystem geschehen. Die zu untersuchenden Medien müssen zuerst aus Quellen wie Social Networks oder Nachrichtenseiten gewonnen werden. Die gewonnenen Daten müssen datenschutzkonform behandelt werden, was beispielsweise eine sichere Ablagestrategie erfordert. Und die Daten müssen bereinigt werden, damit die eigentlichen Analysewerkzeuge nicht durch Rauschen gestört werden.

Bevor in den folgenden Abschnitten der Stand der Forschung und die eigenen Ergebnisse diskutiert werden, sollen zuerst noch einige Begriffe und ihre Verwendung in der Technik eingeführt werden. Die Echtheit eines Inhalts hat in der IT Sicherheit mehrere Aspekte. Eine Frage ist, ob das Dokument und sein Erzeuger die sind, die sie vorgeben zu sein. Dies ist die Frage der Authentizität. Ein Foto, welches von einem bekannten Pressefotografen stammt, wird ein hohes Vertrauen genießen. Es kann aber sein, dass entweder der Fotograf das Foto nicht erstellt hat, sondern seine Urheberschaft nur angegeben wurde, oder dass das Foto selbst ausgetauscht wurde. Ein Beispiel: In einer Kriegsberichtserstattung wird ein Foto verwendet, welches angeblich von dem bekannten (fiktiven) Fotografen Herrn Max Mustermann geschossen wurde. Es zeigt vorrückende Panzer. Die Authentizität stellt zum einen die Frage, ob Herr Mustermann wirklich der Ersteller des Fotos ist. Sie hinterfragt in dem Fall, dass Herr Mustermann tatsächlich Fotos aus dem Kriegsbericht erstellt hat, auch, ob genau dieses Foto auch tatsächlich aus der Menge der Fotos stammt. Weiterhin stellt sich auch die Frage der Integrität. Es ist denkbar, dass ein Foto verändert wurde, um seinen Inhalt zu manipulieren. Dabei ist es unwichtig, ob dies von einer dritten Partei oder dem ursprünglichen Fotografen stammt. Erst, wenn sowohl Authentizität als auch Integrität belegt sind, ist ein Dokument vertrauenswürdig.

A. Überblick über die Forschungslandschaft

Der Bedarf nach einer computergestützten Erkennung von Desinformation ist nicht neu. Insbesondere im englischsprachigen Raum ist hier, seitdem das Phänomen breite Beachtung erfährt, eine Reihe von Forschungsarbeiten durchgeführt worden. Unterschieden werden kann hierbei zwischen Ansät-

zen auf textueller Ebene und auf Basis von Metadaten. Erstere verwenden computerlinguistische Methoden, um Texte mit entsprechendem Inhalt zu erkennen. Letztere nutzen Informationen wie die Aktivitäten von Benutzerkonten oder IP-Adressen, um Bot-Netze zu identifizieren, welche zur Verbreitung eingesetzt werden.

Darüber hinaus sind aber auch Arbeiten aus der Multimedia-Forensik von Bedeutung. Auch die manipulierende Verwendung von Bildern und Videos wird betrachtet, daher sind auch Arbeiten zur Erkennung entsprechender Manipulationen von Bedeutung.

I. Erkennung von Desinformationen bei Texten

Es existieren Webseiten, die in der Vergangenheit aufgefallen sind, Desinformationen (im Folgenden auch „Fake News“ genannt) verbreitet zu haben. Im englischsprachigen Raum gehören zu solchen Webseiten beispielsweise denverguardian.com, wtoe5news.com oder ABCnews.com.co. Oft sind solche Webseiten professionell erstellt und kaum von großen Mainstream-Medienwebseiten zu unterscheiden. Der Hauptzweck der Fake-News-Webseiten ist allerdings die Verbreitung politischer Propaganda oder Profitgenerierung durch die Verbreitung von Desinformationen (beispielsweise durch „Clickbaiting“). Neben den Nachrichtentypen (Propaganda und Clickbaiting) existieren weitere Genres, welche an der Verbreitung von Fake News beteiligt sind z. B. Satire, Verschwörungstheorien, Hoax, „Promi-News“ oder Falschaussagen (z. B. von Politikern) (Rashkin et al., 2017; Rubin et al., 2015).

Allcott et al. (2017) auf der anderen Seite schließen 1) Fehlermeldungen, 2) unbestätigte Gerüchte über Menschen, Ereignisse oder Organisationen, 3) Verschwörungstheorien, 4) Satire, 5) falsche Aussagen von Politikern und 6) Artikel, die irreführend, aber nicht völlig falsch sind, aus dem Konzept von Fake News aus. Die folgende Grafik zeigt unterschiedliche Fake-News-Typen und wie diese ihrem Wahrheitsgehalt und ihrer Intention zugeordnet werden können.

Neben unterschiedlichen Typen existieren auch unterschiedliche Medienträger und Internet-Plattformen (z. B. Webseiten, Social-Networks-Kanäle, Instant-Messaging-Dienste etc.), welche zur Verbreitung von Fake News benutzt werden. An der Verbreitung können Menschen, aber auch sogenannte Bots oder Cyborgs beteiligt sein. Bots sind Computerprogramme, welche überwiegend in Social Networks automatisch Inhalte verbreiten. Cyborgs

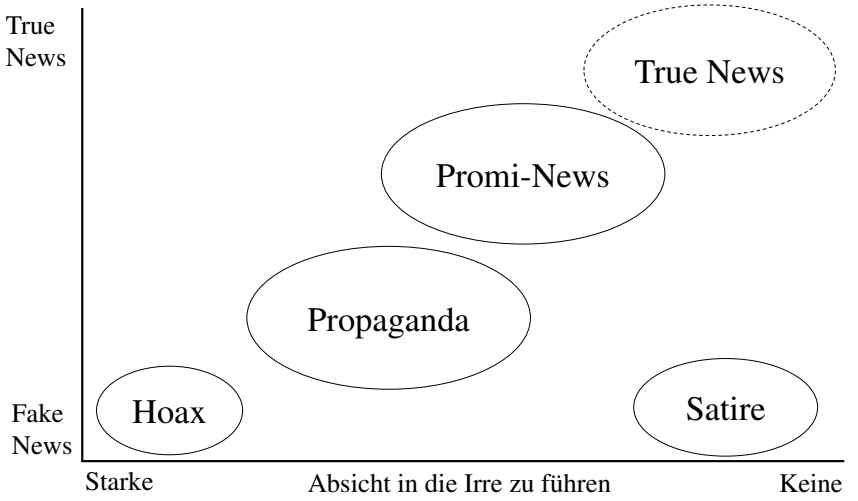

Abbildung 4.1: Nachrichtentypen angelehnt an Rashkin et al. (2017) und Rubin et al. (2015)

sind Accounts, welche zwar von Menschen betrieben werden, deren Inhaltsverbreitung allerdings automatisch abläuft.

Für wissenschaftliche Analysezwecke können sowohl physische, als auch Metadaten einer Meldung verwendet werden. Ein Nachrichtentext verfügt beispielsweise über einen Titel, Textkörper, Veröffentlichungsdatum oder auch zur Nachricht gehörende Fotos und Videos. Als nicht-physisch kann die Semantik des Textes zu Untersuchungszwecken herangezogen werden, beispielsweise der Inhalt der Nachricht, Emotionen, Stimmungen oder das Thema. Alle Eigenschaften (Features) der Nachricht können verwendet werden, um Nachrichtentexte ihrem Wahrheitsgehalt nach zu klassifizieren. Dabei werden die folgenden drei Analyseverfahren unterschieden: linguistic and semantic-based analysis, knowledge-based analysis und style-based analysis.

1. Knowledge-based Analysis

Beim wissensbasierten Ansatz wird der Wahrheitsgehalt eines Nachrichtenartikels direkt (manuell oder maschinell) geprüft (Shu et al., 2017). Bei der manuellen Überprüfung wird der Inhalt des Artikels von Experten z. B. Fachjournalisten überprüft, indem Fakten aus unterschiedlichen Quellen

verglichen werden (Banko et al., 2008). Im englischsprachigen Raum existieren unzählige „Fact-Checking"-Webseiten wie Snopes.com, PolitiFact.com und FactCheck.org. Die Inhalte der Nachrichtenartikel werden von Experten nach dem Wahrheitsgehalt der Behauptungen im Text klassifiziert. Solche Datenquellen sind essentiell für maschinelle Lernverfahren, da sie Trainingsdatenmaterial für den Algorithmus liefern.

Für den wissensbasierten Ansatz werden ebenfalls Crowdsourcing-Ressourcen verwendet. Dabei werden verdächtige Inhalte gemeldet und anschließend von Redakteuren geprüft. Auf diese Weise wird nicht jeder Text überprüft, sondern nur Texte, welche von Nutzenden gemeldet wurden, was einen geringeren Arbeitsaufwand bedeutet.

Neben dem manuellen und Crowdsourcing-Ansatz existieren auch automatische Systeme und Tools, welche Nachrichtenartikel aus dem Internet extrahieren und den Inhalt mit Informationen einer Wissensdatenbank abgleichen. Bestehen Widersprüche zwischen den Texten, wird dies vom System gemeldet (Shu et al., 2017; Banko et al., 2008). Pan et al. (2018) haben unterschiedliche automatische computerbasierte Ansätze zur Erkennung von Fake News getestet und konnten eine Gesamterkennungsrate von etwa 80 % erreichen. Dennoch adressieren die Autoren einige technische Herausforderungen und merken an, dass die rechenorientierte Faktenprüfung nicht umfassend genug ist, um alle Zusammenhänge abzudecken, die für die Erkennung gefälschter Nachrichten notwendig sind.

2. Style-based Analysis

Stilbasierte Ansätze zielen darauf ab, desinformierende Nachrichten anhand des Schreibstils des Verfassers zu erkennen. Hierfür werden syntaktische und semantische Textmerkmale wie etwa die Rechtschreibung, Grammatik, Interpunktion, Wortwahl, Satzbau, Absatzstruktur analysiert. Geeignete Merkmale, um den Schreibstil in Social Networks zu bestimmen sind Tokens wie etwa die Anzahl der URLs, Hashtags, @-Accountnennungen und die Verwendung von Großbuchstaben (Castillo et al., 2011; Horne & Adali 2017; Jin et al., 2016).

Es ist bekannt, dass Schöpfer von Falschnachrichten ihren Schreibstil versuchen zu verschleiern, um nicht als Autor identifiziert zu werden. Es werden aber auch Maschinen trainiert, welche automatisch Falschnachrichten generieren (Social Bots). Oder es werden Inhalte von Webseiten plagiiert, um

diese auf der eigenen Webseite anzubieten und auf diese Weise Einnahmen zu generieren (Potthast et al., 2016).

Um Verfasser von Fake News anhand ihrer Schreibstile zu überführen, haben Potthast et al. (2018) das Unmasking-Verfahren angewendet und auf diese Weise die Stilähnlichkeit zwischen Mainstream-Nachrichtentexten, Satire sowie extrem rechts- und linksseitigen Texten bestimmt. Unmasking entfernt iterativ die stärksten und schwächsten Features, sodass die Texte nach jeder Iteration anhand der übrig geblieben Merkmale verglichen werden. Mit den stärksten und schwächsten Merkmalen sind solche gemeint, die zur Unterscheidung der Texte am meisten bzw. wenigsten beitragen. Das Resultat sind sogenannte Degradationskurven, wobei jeder Text-zu-Text-Vergleich genau eine solche Degradationskurve darstellt. Kurven die stark abfallen bedeuten, dass die Texte nach der iterativen Entfernung von Merkmalen wesentlich schlechter unterschieden werden können (und dementsprechend vom selben Autor stammen) als Kurven, die kaum bis gar nicht abfallen.

Die extrem einseitig verfassten Texte („hyperpartisan"), weisen nach der Unmasking-Methode einen ähnlichen Schreibstil auf und können von Mainstream-Texten mit einer Genauigkeit von 78 % unterschieden werden. Satire kann zu 81 % richtig bestimmt werden. Shu et al. (2017) haben ebenfalls in ihrer Untersuchung gezeigt, dass Satire sowohl von extrem links- und rechtsseitigen Nachrichtentexten als auch Mainstream-Artikeln diskriminiert werden kann.

In der Forschung wurden unterschiedliche maschinelle Lernverfahren angewendet, um Fake News zu erkennen. Zu den klassischen Verfahren gehören beispielsweise Entscheidungsbäume (Decision Trees), Support Vector Machines (SVM), die Logistische Regression und der Nächste-Nachbarn-Klassifikator (KNN) (Afroz et al., 2012; Castillo et al., 2011; Davis et al., 2016; Horne & Adali, 2017; Kwon et al., 2013; Tacchini et al., 2017; Yang & Counts, 2010). Neben den unterschiedlichen Lernverfahren wurden ebenfalls unterschiedliche Datensets und Textmerkmale verwendet, um Fake News zu erkennen. Ebenfalls Anwendung finden „Deep Learning Algorithmen", welche nicht auf explizit definierte Textmerkmale angewiesen sind und Nachrichtenkontextinformationen selbständig erlernen.

Überwachte Lernverfahren hängen stark von der Qualität des Datensatzes ab. Aus folgenden Gründen ist es jedoch schwierig, einen qualitativ hochwertigen Datensatz zur Erkennung von Falschnachrichten zu generieren: Die Daten im Internet sind oft unstrukturiert, verrauscht und unvollständig (Shu et al., 2017). Jeden Tag erhöht sich die Menge an Falschinformationen, welche mit unterschiedlichen Absichten verfasst werden. Unüberwachte Lernmodel-

le sind daher praktikable Lösungen für solche Probleme aus der Praxis. Es gibt jedoch nur wenige Studien, die unüberwachte Lernverfahren zur Erkennung von Fake News anwenden. Die meisten von ihnen konzentrieren sich auf semantische Ähnlichkeits- oder Sentimentanalysen. Ahmed (2017) schlägt in seiner Dissertation ein unüberwachtes Ähnlichkeitsmessverfahren vor zur Erkennung von gefälschten Rezensionen. Für sein Verfahren verwendete er Wort- und Wortfolgenähnlichkeiten als Textmerkmale und konnte auf diese Weise ähnliche Rezensionen bzw. Duplikate mit einer hohen Erkennungsrate klassifizieren. Dieses Verfahren könnte künftig für die Erkennung von Falschinformationen eingesetzt werden, da Nachrichten von Qualitätsmedien oft plagiiert und in leicht abgewandelter Form für bestimmte politische oder soziale Ziele missbraucht werden.

Abschließend lässt sich festhalten, dass viele Techniken zur automatischen Erkennung gefälschter Nachrichten vorgeschlagen und angewendet werden. Die Faktenüberprüfung obliegt nichtsdestotrotz immer noch dem Wissen der Menschen.

II. Erkennung von Bildmanipulationen

Die Identifizierung von Bildmanipulationen ist ein komplexes Feld, insbesondere, wenn man auch die Wiederverwendung von Bildern bereits als potenziellen Angriff betrachtet. Es existieren aber auch zahlreichen Methoden zur Bildmanipulation, die die Bedeutung eines Bildes verändern. Hinzu kommt die Wiederverwendung von Bildern, die nicht verändert, sondern in einem anderen Kontext verwendet werden. Bildmontagen sind eine Kombination aus beiden Ansätzen: Ein bestehendes Bild wird außerhalb des Kontextes verwendet, aber auch manipuliert, indem ein weiteres Objekt aus anderen Bildern hinzugefügt wird, um seine Bedeutung zu ändern.

Es gibt viele Methoden zur Erkennung von Manipulationen in digitalen Bildern. Heute lassen sie sich grob zwei Klassen zuordnen: Methoden, die auf vom Menschen definierten Modellen und Mustern basieren, und Methoden, die auf überwachtem maschinellen Lernen basieren. Für die erste Gruppe sind bereits mehrere Veröffentlichungen, die einen Überblick zum Thema bieten, beispielsweise von Bayram (2008) oder Birajdar (2013), verfügbar.

Ein generischer Ansatz besteht darin, Unterschiede in den Stärken von Fehlern zu berechnen, die durch unterschiedliche Kompressionshistorien von Bereichen verursacht werden, wie beispielsweise von Luo (2010) beschrieben.

Die Re-Identifikation von Bildern ist eine Domäne, in der robuste Hashverfahren (Shin, 2013) oder wahrnehmungsbasierte Hashverfahren (Niu, 2008) die besten Ergebnisse liefern.

Die Herausforderung bei der Re-Identifikation besteht darin, Bilder zu identifizieren, die Kopien einer bekannten Quelle sind, ohne zu empfindlich auf die Ablehnung von Bildänderungen durch übliche Bildverarbeitung, z. B. verlustbehaftete Kompression, zu reagieren. Andererseits muss man es vermeiden, ähnliche Bilder als Duplikate zu identifizieren, da die Ähnlichkeit nicht für eine tatsächliche Anerkennung der Wiederverwendung qualifiziert.

1. Manipulationserkennung durch maschinelles Lernen

Es gibt zwei Arten von Manipulationen an digitalen Bildern: zufällige Manipulationen und absichtliche Manipulationen. Ersteres beinhaltet die durch die gängige Bildverarbeitung verursachten Änderungen wie Rauschfilterung, Skalierung, Komprimierung und so weiter. Letzteres bezieht sich in der Regel auf Bildfälschungen, die auf Inhaltsänderungen wie das Einfügen und Entfernen von Objekten abzielen. Ziel der Bildverarbeitung ist es nicht, den Bildinhalt zu verändern, sondern die Bildqualität oder Speichereffizienz zu verbessern. Die Bildverarbeitung ändert nur die binäre Darstellung des Bildes, während die zugrundeliegende semantische Bedeutung intakt bleibt. Daher wird die zufällige Manipulation auch als inhaltserhaltende Manipulation bezeichnet. Im Gegensatz dazu beabsichtigt die Bildfälschung, die semantische Bedeutung des Bildes zu bearbeiten, die auch als inhaltsverändernde Manipulation oder böswillige Manipulationen bezeichnet wird.

Bildforensik zielt auf die Erkennung beider Arten von Manipulationen ab. Die Erkennung absichtlicher Manipulationen kann die manipulierten Teile lokalisieren und die Absicht der Fälschung ableiten. Das Erkennen von zufälligen Manipulationen kann den Verarbeitungsverlauf aufdecken und auf mögliche Manipulationen hinweisen. In der Praxis erfolgt in der Regel nach der Bildverfälschung eine weitere Bildverarbeitung, um die Bearbeitungsspuren zu verbergen. Zum Beispiel zeigt das Erkennen von Skalierung und Filterung an, dass ein Bild nach der Aufnahme bearbeitet wurde. Die Existenz der doppelten JPEG-Komprimierung verrät, dass ein Bild nicht die Originalkopie ist.

Klassische bildforensische Techniken versuchen, bestimmte Bildoperationen, wie z. B. Medianfilterung, Skalierung, doppelte JPEG-Komprimierung, Copy-Move, etc., durch Untersuchung der entsprechenden Spuren zu er-

kennen [1-2]. Die Erkennung basiert auf der manuellen Analyse von manipulierten Bildern und extrahierten Merkmalen, deren Eigenschaften die eindeutigen Spuren der Manipulationen zeigen. Die Merkmale sind in der Regel mit der Manipulation verbunden und ein bestimmtes Merkmal wird verwendet, um eine bestimmte Manipulation zu erkennen. Daher ist jede klassische forensische Technik auf eine bestimmte Bildoperation ausgerichtet. Zum Beispiel kann der lokale Rauschpegel verwendet werden, um Image Splicing zu erkennen, indem die Rauschvarianz innerhalb kleinerer Bildblöcke bewertet wird. Die statistischen Merkmale der DCT-Koeffizienten in JPEG werden allgemein angewendet, um doppelte Komprimierung zu erkennen und die Manipulationen zu lokalisieren. Darüber hinaus werden einige allgemeine Low-Level-Merkmale, wie z. B. Pixel-Kookkurrenz, auch zur Unterscheidung verschiedener Bildverarbeitungsoperationen verwendet.

Im Gegensatz dazu verwenden Deep Learning forensische Methoden neuronale Netze, um wichtige Merkmale automatisch aus einem großen Trainingsdatensatz zu lernen [3-5]. Aufgrund seines sehr guten Ergebnisses für die Objekterkennungsaufgabe in digitalen Bildern wird das Convolutional Neural Network in der Bildforensik zur Erkennung von Manipulationen eingesetzt. Eine große Herausforderung für eine effektive Erkennung von Bildmanipulationen stellt die verbreitete JPEG-Komprimierung dar, die sowohl für klassische als auch für tief lernende Ansätze gilt. Die JPEG-Komprimierung verwirft Bildinformationen im Hochfrequenzbereich, was dazu führt, dass die Spuren, die bei Manipulationsoperationen hinterlassen werden, verringert oder zerstört werden, während gleichzeitig neue spezifische Spuren eingeführt werden. Abhängig vom Qualitätsfaktor eliminiert die JPEG-Komprimierung mehr oder weniger Bilddetails.

2. Merkmalserkennung

Die Merkmalserkennung besteht darin, sogenannte Schlüsselpunkte (Keypoints) an mehreren Stellen in einem Bild mit einem Detektor zu erkennen und Deskriptoren mit einem Merkmals-Extraktor zu extrahieren. In einem weiteren Schritt, dem Merkmalsvergleich, werden die gefundenen Merkmale mit Merkmalen eines anderen Bildes verglichen. Wenn beide Bilder nun das gleiche Objekt enthalten, sollten die Merkmale idealerweise messbar ähnlich sein. Ein Merkmal selbst ist definiert als ein „interessanter" Teil des Bildes. Was genau als „interessanter" Teil des Bildes verstanden wird, variiert je nach Merkmalsdetektor. Der Bildteil, in dem ein Merkmal extrahiert wird,

ist oft entweder ein isolierter Punkt, eine kontinuierliche Kurve oder ein verbundener Bereich.

Der Scale Invariant Feature Transform (SIFT) (Lowe, 2004) Algorithmus ist wahrscheinlich einer der bekanntesten und am häufigsten verwendeten Merkmals-Detektoren. SIFT findet Schlüsselpunkte in einem Bild mit dem Difference-of-Gaussian (DOG) Operator. DOG wird verwendet, um über mehrere Bildgrößen nach lokalem Extrema zu suchen. So können Schlüsselpunkte gefunden werden, die auch bei einer Änderung der Bildgröße erhalten bleiben. Für jeden dieser Schlüsselpunkte wird dann die stärkste Orientierung aus der Nachbarschaft zugewiesen.

Der Speeded Up Robust Features (SURF) Detektor (Bay, 2006) ist teilweise von SIFT inspiriert und ist ein Versuch, schneller und robuster zu sein als SIFT. Wie SIFT basiert auch SURF auf dem Difference-of-Gaussian (DOG). Um Schlüsselpunkte zu erkennen, verwendet SURF eine ganzzahlige Approximation eines Bildobjekt-Detektors.

3. Erkennung von Deepfake-Videos

Eine besondere Art der Manipulation zielt in Bildern und, vor allem, in Videos auf darin abgebildete Personen: Mit der sog. Deepfake-Technologie kann das Gesicht einer Person A automatisiert durch das Gesicht einer anderen Person B beinahe „lebensecht" ersetzt werden: Mimik und Mundbewegungen, die Kopfhaltung und auch das passende Größenverhältnis und die perspektivische Ansicht werden für das ausgetauschte Gesicht in jedem Einzelbild des Videos automatisch eingepasst.

Hiermit lassen sich innerhalb weniger Stunden bis Tage gefälschte Videos erzeugen, die die (Ziel-) Person B scheinbar in einer kompromittierenden Situation zeigen, indem ihr Gesicht in u. U. brisantes Videomaterial nachträglich eingefügt wird (etwa in Pornografie).

Die Technologie ist frei verfügbar, beispielsweise als kostenfreie Software von Kowalski (2018), deepfakes (2019), shaoanlu/clarle (2019), als „Deepfakes FakeApp" aus diversen Internetquellen oder – noch bequemer – als Online-Dienstleistung bei Deepfakes web β (2019).

Der Name „Deepfake" leitet sich davon ab, dass Techniken des sog. Deeplearnings aus dem Gebiet des maschinellen Lernens genutzt werden.

Auch Verfahren zur Erkennung dieser Deepfake-Videos sind bekannt: Einige basieren darauf, dass die Bildbereiche der gefälschten Gesichter einen anderen „technischen Lebenszyklus" durchlaufen haben als die unveränder-

ten Bildbereiche des Bildhintergrundes. Das betrifft die Anzahl und Parameter der Video- und Bildkompressionen (Bianchi & Piva, 2012) oder die Vergrößerung und/oder Rotation der eingepassten Bildinhalte (Li, Yuan et al., 2009), welche in den echten und gefälschten Bildbereichen voneinander abweichende forensische Spuren hinterlassen.

Andere Verfahren können ggf. eine unnatürliche Häufigkeit des Augenblinzelns der Personen im Video erkennen, so etwa von Li, Chang et al., 2018.

Einen generalisierten Ansatz verfolgen Arbeiten, die ihrerseits – ebenso wie der Deepfake-Algoritmus – maschinelles Lernen einsetzen, so etwa von Rössler, Cozzolino et al. (2018), Afchar, Nozcik et al. (2018) und Chollet (2016).

III. Bot-Erkennung

Bei der Bot-Erkennung geht es um die Frage, ob die Aktivitäten von Nutzerprofilen einer Online-Plattform von Menschen stammen oder ob sie programmgesteuert sind. Wenn hinter den Interaktionen von Profilen ein Computerprogramm steckt (wenn auch nur zeitweise), handelt es sich um ein (halb-)automatisches Profil bzw. um einen (Social) Bot. Der transparente Einsatz von Bots kann viele positive Effekte haben, beispielsweise hat er ein Rationalisierungspotenzial für Anbieter und auf Kundenseite kann er zu einem Gewinn an Gebrauchstauglichkeit führen.

In Zusammenhang mit Desinformation werden Bots jedoch in intransparenter Weise zur Verstärkung der Verbreitung von Fake News eingesetzt, hier kommen „Malicious Social Bots“ zum Einsatz. Das an „one man, one vote“ angelehnte Prinzip „one man, one voice“ wird durch diese Bots gebrochen: Das Ziel ihres Einsatzes ist die Manipulation von Meinungen mittels Fake News, die scheinbar von einer Masse von Menschen für wahr gehalten und verbreitet wird.

Zur Bot-Erkennung werden verschiedene Verfahren erforscht. Sie lassen sich grob bezüglich ihrer jeweiligen Methodik (s. Karataş & Şahin, 2017) in drei Kategorien einteilen (wobei zur Detektion vielfach Kombinationsverfahren hiervon eingesetzt werden):

– Bot-Erkennung mittels Methoden der Strukturerkennung,
– Bot-Erkennung durch Crowdsourcing und
– Bot-Erkennung mit Methoden des maschinellen Lernens.

Bacciu et al. (2019) erzielen auf Twitter eine Bot-Erkennung von ca. 95 % bei englischsprachigen Texten.

Es ist offensichtlich, dass die Detektionsergebnisse durch die Betreiber wesentlich besser sind, als diejenigen, die von außerhalb des Netzes erzielt werden können. Die Betreiber haben – im Gegensatz zu dritten Parteien – globalen Vollzugriff mit beliebigen, eigenen Programmierschnittstellen (API = application programming interface) auf die Primärdaten der Profile, sie können wie keine andere Organisation Big-Data-Analysen (unter Wahrung des Datenschutzes) hierauf durchführen und in Echtzeit im Direktzugriff bei aktuell aktiven Profilen die Provenance der Aktivitäten analysieren.

Aufgrund der besonderen Bedeutung von Twitter für Nachrichtenmedien, der vielfältigen Möglichkeiten des Einsatzes von APIs bei Twitter und weil es bei Twitter keine Echtnamen-Pflicht gibt, werden im Folgenden Analyseergebnisse auf Twitter mittels maschinellen Lernens zusammengefasst.

Gilani et al. (2017) analysierten Verhaltensmerkmale von Bots und Menschen in Twitter-Daten, wie z. B. Vorlieben, Retweets, Antworten und @-Mentions, Aktivität oder die Menge der hochgeladenen Inhalte. Sie zählten auch die gemeinsamen URLs und das Verhältnis von Followern für jedes Twitter-Benutzerkonto. Für die Datenerfassung, Vorverarbeitung, Annotation und Analyse wurde das als Stweeler (Gilani et al., 2016) bekannte Framework verwendet.

Die Autoren von Gilani et al. (2017) beobachteten, dass Menschen neuere Inhalte generieren, während Bots auf Retweeting angewiesen sind. Außerdem haben Bots eine höhere Neigung, URLs zu teilen und Medien (wie Bilder und Videos) häufiger hochzuladen als Menschen. Varol et al. (2017) stellten ein Twitter-Bot-Erkennungsframework vor, das mehr als tausend Features aus sechs verschiedenen Klassen von Twitter-Benutzerdaten und Metadaten extrahiert. Extrahierte Features sind z. B. die Anzahl der Freunde des Twitter-Nutzers, der getwitterte Inhalt, die Stimmung im Text oder seine Aktivitätszeitreihe. Die extrahierten Merkmale wurden dann verwendet, um verschiedene Modelle zur Boterkennung zu trainieren. Durch ein 5-faches Kreuzvalidierungsverfahren erreichte der trainierte Random Forest Klassifikator eine Genauigkeit von 0,95 AUC.

B. Überblick über die eigene Forschung

Im Projekt DORIAN wurden zahlreiche Methoden untersucht, wie Desinformationen erkannt werden können. Dabei war die Absicht nicht, fertige

Werkzeuge zu entwickeln, mit denen bereits automatisiert Inhalte beurteilt werden können. Es galt vielmehr, die vielfältigen Möglichkeiten dieser Erkennung zu sondieren und sie im Rahmen erster Testimplementierungen auf ihre Eignung hin zu untersuchen. Dabei wurde auch ein Augenmerk auf die Gestaltung einer passenden Umgebung gelegt, die einen effizienten Zugriff auf Interneträume durch Crawler ermöglicht und dabei die Prinzipien des Privacy-by-Design (Cavoukian, 2009) beachtet.

I. Datenerfassung mittels Crawling-Technologie

Für die automatisierte Erfassung von Informationen aus verschiedenen Interneträumen wurde im Projekt DORIAN ein Framework konzipiert, bei dem der Einsatz von Crawling-Technologie im Fokus steht. Für einzelne Module des Frameworks wurden Testimplementierungen durchgeführt und evaluiert.

1. Definitionen

Crawler

Als „Webcrawler“ oder kurz „Crawler“ werden Programme bezeichnet, die sich automatisiert meist über Verlinkungen (URLs) zwischen Webseiten durch das Internet bewegen, um die auf den verschiedenen Webseiten vorhandenen Inhalte zu erfassen. Ein Crawler stellt dabei grundsätzlich zwei Mechanismen zur Verfügung: Einen Mechanismus zur Fortbewegung und Datenerfassung im Internet, also das Absetzen von Anfragen über das Hypertext Transfer Protocol (HTTP) und das Herunterladen der entsprechenden HTTP-Antworten, sowie einen Mechanismus zur Analyse der erfassten Inhalte, beispielsweise um die URL für die nächste Webseite zu ermitteln. Die Inhalte, die ein Crawler erfassen/herunterladen kann, sind dabei nicht nur auf HTML-Webseiten beschränkt, sondern schließen jegliche Dateitypen (bspw. Dokumente, Bilder, Videos, …) ein, solange diese über HTTP übertragen werden.

Scraper

Diejenige Teilkomponente eines Crawlers, die die Analyse der erfassten Inhalte übernimmt, wird nachfolgend „Webscraper" oder kurz „Scraper" bezeichnet. Scrapers sind zentrale Bestandteile eines Crawlers: Sie ermöglichen eine Selektion derjenigen Inhalte oder Teilbereiche von Webseiten, die für die jeweilige Datenerfassung relevant sind. Entsprechend kann hier bereits eine Filterung oder Säuberung größerer Datenmengen stattfinden, bspw. wenn eine Vielzahl von Webseiten erfasst, auf jeder Webseite aber lediglich die enthaltenen URLs gesammelt und alle anderen Inhalte ignoriert werden sollen. Scrapers analysieren und strukturieren dazu die erfassten HTML-Webseiten, um mittels geeigneter Abfragesprachen wie XPath die Selektion relevanter Inhaltsteile zu ermöglichen. Hier wird deutlich, dass ein Scraper stets webseitenspezifisch ist: Zwei im Aufbau grundsätzlich verschiedene Webseiten benötigen jeweils einen eigenen Scraper, da sich ihre Struktur und somit ebenfalls die Selektion der relevanten Inhalte unterscheidet. Entsprechend hoch ist der Aufwand der Entwicklung und Wartung eines Crawlers, der eine Vielzahl verschiedener Webseiten automatisiert bearbeiten können soll.

Dynamische Inhalte

Typischerweise verarbeitet ein einfacher Crawler mittels seiner integrierten Scrapers Webseiten in ihrer „rohen" Form: Der Crawler lädt die Webseite als HTML herunter und übergibt diese seinem zuständigen Scraper zur Analyse und Strukturierung. Hierbei werden typischerweise eingebundene Inhalte wie bspw. Bilder, Videos, Stylesheets oder JavaScript-Dateien nicht weiter betrachtet. Dies hat zur Folge, dass insbesondere über JavaScript nachgeladene dynamische Inhalte nicht vom Scraper analysiert werden können – der Scraper verarbeitet entsprechend nur unvollständige Daten und liefert im schlimmsten Fall falsche Ergebnisse, bspw. in Form einer leeren Webseitenstruktur.

Um auch dynamische Inhalte in korrekter Form erfassen und verarbeiten zu können, ist es notwendig, Scraper mittels zusätzlicher Mechanismen zu erweitern. Hierbei bieten sich Test-Frameworks wie Selenium an: Dieses Framework ist eigentlich für Softwaretests von Webanwendungen entwickelt worden und ermöglicht quasi die automatisierte Steuerung eines vollständigen Webbrowsers. Hiermit wird sichergestellt, dass dynamische Inhalte nachgeladen und ggf. notwendige Interaktionen auf der Webseite (bspw.

Button-Klicks, Scrolling, etc.) durchgeführt werden können, um die Inhalte der Webseite vollständig zu erfassen. Die Integration von Selenium in eigene Scrapers ist für eine Vielzahl von Programmiersprachen möglich und einfach umzusetzen, hat jedoch zum Nachteil, dass die jeweiligen Scrapers mehr Ressourcen benötigen und entsprechend schwerfälliger ihre Prozesse abarbeiten.

Application Programming Interfaces (APIs)

Eine Application Programming Interface (API) im Web-Bereich stellt eine Schnittstelle dar, die von einem Inhaltsanbieter externen Entwicklern zur Verfügung gestellt wird, damit diese über (ebenfalls externe) Programme effektiv und effizient vordefinierte Prozesse auf den Inhalten des Anbieters ausführen können, bspw. Suchen im Datenbestand oder die Steuerung konkreter Verarbeitungsschritte. APIs sind typischerweise zugriffsgeschützt und benötigen einen entsprechenden Account beim Anbieter, dem die Nutzung der API gestattet ist. Zu seiner API liefert der Anbieter typischerweise eine umfangreiche Spezifikation und Dokumentation und legt außerdem die Kriterien für die Nutzung seiner API fest, bspw. Limitierungen der jeweiligen Abfragen, Datennutzungsbestimmungen oder Anforderungen für den Erhalt eines Zugangs zur API.

Aus Sicht eines Entwicklers ist der Zugriff auf die API eines Anbieters relevanter Inhalte stets zu bevorzugen, da typischerweise der Programmieraufwand durch die existierende API-Spezifikation deutlich reduziert wird. Jedoch können die Anforderungen eines Anbieters für die Nutzung seiner API durchaus so hoch sein, dass eine API-Nutzung aus technischer oder auch finanzieller Sicht impraktikabel wird.

2. Crawling-Framework für größerer Interneträume

Im Rahmen der Projektarbeit wurde ein Konzept für ein Framework entworfen, das in der Lage ist, in größeren Interneträumen Daten zu erfassen, potenziell relevante Inhalte zu selektieren und abzuspeichern. Die so erfassten Daten sollen eine Grundlage für nachfolgende Analysen hinsichtlich Desinformationen darstellen. Als Interneträume wurden exemplarisch drei große Plattformen (YouTube, Twitter und Facebook) betrachtet, auf denen Nutzende zahlreiche eigene Inhalte veröffentlichen und die Inhalte anderer

kommentieren können. Zur Datenerfassung wurden Testimplementierungen durchgeführt und evaluiert.

Das Crawling-Framework ist modular konzipiert, wobei verschiedene Scraping-Module für jeweils verschiedene Scraping-Tasks eingesetzt werden. Für die Speicherung der erfassten Daten wird ein relationales Datenbanksystem verwendet. Als relevante Inhalte werden auf den drei Plattformen jegliche Textinhalte betrachtet, die von Nutzenden erstellt worden waren. Das entspricht auf YouTube den Kommentaren unter den jeweiligen Videos, bei Twitter den Tweets und bei Facebook den Posts und Kommentaren der Nutzerinnen und Nutzer. Alle drei Pattformen unterscheiden sich in ihrer Struktur und anhand ihrer Inhalte signifikant, sodass jeweils ein Scraping-Modul je Plattform vorgesehen ist.

Nachfolgend werden die drei Plattformen einzeln betrachtet und die Möglichkeiten und Herausforderungen einer automatisierten Datenerfassung aufgezeigt.

YouTube-Modul

YouTube bietet über die YouTube-Data-API effizienten Zugriff auf Kommentare samt umfangreicher Metadaten zu Kommentaren, Videos und Nutzenden. Für einen API-Zugang wird lediglich ein Google-Account benötigt. Zur Datenerfassung auf YouTube eignet sich somit ein einfacher Scraper, der Anfragen an die API stellen kann. Zusätzliche Funktionalität für die Verarbeitung dynamischer Inhalte wird entsprechend nicht benötigt.

Es existieren zwar Limitierungen für die Nutzung der API, diese sind jedoch relativ großzügig bemessen: Pro Tag besitzt jeder API-Nutzende ein Kontingent von 10.000 Einheiten, wobei einzelne Abfragen zwischen 2-5 Einheiten verbrauchen.

Zusammenfassung

Relevante Daten auf YouTube lassen sich somit sehr einfach erfassen, eine API-Zugriffsberechtigung lässt sich schnell erlangen und der Implementierungsaufwand eines Scrapers ist gering.

Facebook-Modul

Wie YouTube und Twitter stellt auch Facebook seine Graph API für den Datenzugriff zur Verfügung, für den ein Facebook-Account mit entsprechender API-Berechtigung benötigt wird. Die Graph API von Facebook bietet zwar ebenfalls umfassenden Zugriff auf Nutzer- und Metadaten, jedoch werden selbst nach erteilter API-Zugriffsberechtigung weitere Berechtigungen von Facebook eingefordert, die die Nutzung der API quasi gänzlich impraktikabel machen: Facebook fordert, dass Nutzende der API, die Daten über Facebook-Nutzende erfassen möchten, von jedem dieser Facebook-Nutzende eine Einwilligung über diese Datenerfassung einholt. Dies kann automatisiert über eine Abfrage auf Seite des Facebook-Nutzenden geschehen und wird bspw. bei der Installation von Apps für Smartphones eingesetzt, die eine Facebook-Integration anbieten. Willigt der Facebook-Nutzende ein, so kann die entsprechende App bzw. das Unternehmen, das die App zur Verfügung stellt, die Daten des Facebook-Nutzenden über die Graph API erfassen. Ohne Einwilligung ist dies jedoch nicht möglich. Somit ist die automatisierte Erfassung von relevanten Daten auf Facebook nicht über die Graph API realisierbar, solange kein Weg gefunden wird, von sämtlichen Facebook-Nutzenden, die relevante Inhalte produzieren, eine Einwilligung zu erhalten.

Entsprechend wird für die Entwicklung eines Scrapers für Facebook die Betrachtung der Facebook-Webseite notwendig. Ein einfacher Scraper stößt hier jedoch schnell an seine Grenzen: Da von Facebook auf seiner Webseite sehr viele dynamische Inhalte eingesetzt werden (insbesondere zum Nachladen von Inhalten durch Paging, Scrolling), ist der zusätzliche Einsatz von Selenium notwendig. Hiermit lassen sich die Struktur der Facebook-Webseite analysieren und einzelne Elemente gezielt ansteuern. Da mit Selenium quasi ein vollständiger Webbrowser automatisiert gesteuert wird, leidet die Leistungsfähigkeit eines Scrapers, der Selenium einsetzt, sehr stark im Vergleich zu einem Scraper, der lediglich einer API-Spezifikation folgen muss und sehr effizient entsprechende Abfragen stellen kann.

Zusammenfassung

Relevante Daten auf Facebook sind über die Graph API quasi nicht erfassbar, da Facebook hier sehr einschränkende Bedingungen stellt. Scraper, die die Graph API umgehen und stattdessen über die Webseite von Facebook Daten

erfassen wollen, müssen zwangsläufig Frameworks wie Selenium einsetzen, wodurch der Scraper vergleichsweise nur sehr langsam Daten erfassen kann. Der Implementierungsaufwand ist verhältnismäßig hoch, da eine Vielzahl von Elementen der Facebook-Webseite angesteuert werden müssen, um die dynamischen Inhalte nachladen zu können.

3. Evaluation von Testimplementierungen

Im Rahmen des Projekts wurden erste Testimplementierungen für jedes Scraping-Modul der drei großen Plattformen nach dem oben beschriebenen Konzept durchgeführt und evaluiert. Hierbei konnte die Funktionsfähigkeit jedes Moduls bestätigt werden, wobei sich teilweise erhebliche Leistungsunterschiede hinsichtlich der Scraping-Dauer zeigten. Während pro Sekunde auf YouTube fast 90 Kommentare und auf Twitter bis zu 1600 vollständige Tweets erfasst werden konnten, dauerte die Erfassung eines einzelnen Kommentars auf Facebook bis zu 3 Sekunden. Hier zeigte sich deutlich der Overhead eines Scrapers, der (im Gegensatz zu Scrapers, die auf eine API zugreifen können) einen vollständigen Webbrowser steuern und umständlich auf Webseiten mit dynamischen Inhalten navigieren muss, um an die gewünschten Inhalte zu gelangen.

II. Datenschutzrechtliche Aspekte der automatischen Erkennung von Desinformationen für wissenschaftliche Forschungszwecke

Im Rahmen der automatischen Erkennung von Desinformationen werden öffentlich gemachte Artikel und Meldungen in Online-Nachrichtenplattformen sowie Beiträge in Social Networks erhoben und analysiert. Die verarbeiteten Daten enthalten in der Regel personenbezogene Daten[1] (z. B. Name des Autors, Inhalte, Abbildungen, Zitate einer natürlichen Person), bei deren Verarbeitung die Anforderungen der Datenschutz-Grundverordnung (nach-

1 Personenbezogene Daten sind diejenigen Informationen, die eine natürliche Person („betroffene Person“) direkt oder indirekt identifizieren oder identifizierbar machen (Art. 4 Nr. 1 DSGVO). Z. B. Vor- und Nachname, die Telefon- und Kreditkartennummer sowie die Hobbies und Interessen einer natürlichen Person. Auch pseudonymisierte Daten eröffnen den Anwendungsbereich der Regeln des Datenschutzrechts. Auf anonymisierte Daten, bei denen eine Identifizierung natürlicher Personen nicht (mehr) möglich ist, finden die Anforderungen des Datenschutzrechts jedoch keine Anwendung.

folgend DSGVO) einzuhalten sind. Nachfolgend werden die wichtigsten Anforderungen der DSGVO näher erläutert, die bei der automatischen Erkennung von Desinformationen für Forschungszwecke zu beachten sind.

1. Zulässigkeit der Verarbeitung

Die automatische Sammlung und Auswertung personenbezogener Daten verstößt grundsätzlich gegen das Grundprinzip der DSGVO, wonach die Verarbeitung personenbezogener Daten auf das notwendige Maß zu beschränken ist (Art. 5 Abs. 1 lit. c, Datenminimierung). Daher kann sie nur in Ausnahmefällen zulässig sein.[2]

Eine solche Ausnahme liegt vor, wenn die Verarbeitung personenbezogener Daten für wissenschaftliche Forschungszwecke erforderlich ist (Art. 89 Abs. 1 DSGVO). Unter wissenschaftliche Forschung ist die technologische Entwicklung sowie die Grundlagenforschung, die angewandte Forschung und die privat finanzierte Forschung zu verstehen.[3] Die Verarbeitung für Forschungszwecke wird in der Datenschutz-Grundverordnung sowie in sonstigen anwendbaren Datenschutzregelungen privilegiert[4] (z. B. hinsichtlich der Zweckbindung und Speicherbegrenzung sowie hinsichtlich der Rechte der betroffenen Personen). Grundvoraussetzung ist, dass die forschende Stelle zusätzliche Garantien zum Schutz der Rechte der Freiheiten der betroffenen Personen vorsieht. Mit technischen und organisatorischen Maßnahmen soll sichergestellt werden, dass bei der Durchführung der Verarbeitungstätigkeiten insbesondere das Prinzip der Datenminimierung eingehalten wird. Die Maßnahmen sind bereits bei der Entwicklung der Mechanismen, d. h. vor der tatsächlichen Verarbeitungstätigkeit zur Erkennung von Desinformationen zu treffen (Art. 25 Abs. 1). Zu den Maßnahmen kann die Pseudonymisierung und die Anonymisierung gehören (Art. 89 Abs. 1 Satz 3). Insofern hat die forschende Stelle zunächst zu prüfen, ob der Forschungszweck auch mit anonymisierten Daten zu erreichen ist. Im Rahmen des DORIAN-Projekts hat sich jedoch gezeigt, dass die Verarbeitung mit anonymisierten Daten für die automatische Erkennung von Desinformationen ungeeignet ist. Die

2 Hoeren, Skript Internetrecht, Stand November 2018, S. 492; Schulz, in: Gola, Datenschutz-Grundverordnung 2018, Art. 6 Rn. 257-258.

3 Buchner/Tinnefeld, in: Kühling/Buchner, Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG 2017, Art. 89 Rn. 13.

4 Z. B. Datenschutzgesetze der Länder, Bundesdatenschutzgesetz, Landeskrankenhausgesetze, Meldegesetze.

Verarbeitung personenbezogener Daten ist erforderlich, um u.a. Verbreiter von Desinformationen zu identifizieren sowie Korrelationen zu erkennen. Hier sollte anstelle der Anonymisierung eine Pseudonymisierung der Daten erfolgen. Die pseudonymisierten Daten sollten darüber hinaus verschlüsselt gespeichert und übertragen werden. Ist die Verarbeitung personenbezogener Daten zu einem späteren Zeitpunkt des Projekts nicht mehr erforderlich, sind die Daten unverzüglich zu löschen bzw. zu anonymisieren. Die Weiterverarbeitung der Daten für andere Forschungszwecke sollte ebenfalls ohne Personenbezug erfolgen (Art. 89 Abs. 1 Satz 4).

2. Dokumentationspflichten

Darüber hinaus hat die forschende Stelle ihren Dokumentations- und Nachweispflichten nachzukommen (Art. 5 Abs. 2 DSGVO). Neben einer Verfahrensbeschreibung (Art. 30 Abs. 1 DSGVO) hat die forschende Stelle Störungen in der Sicherheit der Verarbeitung bei der zuständigen Behörde zu melden (Art. 33 Abs. 1 DSGVO) und in den Fällen, in denen ein hohes Risiko für die Rechte und Freiheiten der betroffenen Personen vorliegt, eine Datenschutzfolgenabschätzung durchzuführen (Art. 35 DSGVO).[5] Hier werden die möglichen Risiken analysiert und bewertet und folglich die effektive Gegenmaßnahmen für die bestehenden Risiken ermittelt und implementiert.

3. Informationspflicht und Rechte der betroffenen Personen

Zur Gewährleistung der Transparenz der Verarbeitung regelt die DSGVO Informationspflichten des Verantwortlichen (Art. 13 und 14) sowie Rechte der betroffenen Personen (z. B. Recht auf Auskunft oder Löschung, siehe Art. 15 ff.). Für die Forschung sieht die DSGVO eine Ausnahme der Informationspflicht vor, wenn die Erteilung dieser Information sich als unmöglich erweist oder einen unverhältnismäßigen Aufwand erfordern würde (Art. 14 Abs. 5 lit. b). Bei der automatischen Erhebung und Auswertung personenbezogener Daten für Forschungszwecke wird in der Regel diese Ausnahme

5 Hierzu: Leitlinien zur Datenschutz-Folgenabschätzung (DSFA) und Beantwortung der Frage, ob eine Verarbeitung im Sinne der Verordnung 2016/679 „wahrscheinlich ein hohes Risiko mit sich bringt“, Datenschutzgruppe nach Art. 29, 4.10.2017, 17/DE WP 248 Rev. 01.

Anwendung finden. Darüber hinaus regelt die DSGVO eine Einschränkung des Löschungsrechts der betroffenen Personen, wenn die Verwirklichung des Rechts die Verarbeitung für wissenschaftliche Forschungszwecke unmöglich macht oder ernsthaft beeinträchtigt (Art. 17 Abs. 3 lit. d). Hier ist in jedem Einzelfall zu prüfen, ob die Voraussetzungen vorliegen. Weitere Einschränkungen der Rechte sowie Ausnahmen der Informationspflichten werden in nationalen Gesetzen vorgesehen (z. B. § 27 Abs. 2 BDSG, § 24 Abs. 2 HDSIG.).

4. Verarbeitung besonderer Kategorien von Daten

Bei der automatischen Erkennung von Desinformationen ist eine Verarbeitung von besonderen Kategorien von Daten nicht auszuschließen. Gegenstand der Analyse können beispielsweise Inhalte sein, die Informationen über die politische oder religiöse Überzeugung einer natürlichen Person enthalten. Gemäß Art. 9 DSGVO ist die Verarbeitung dieser Art Daten grundsätzlich verboten. Diese Daten sind als besondere Kategorien von personenbezogenen Daten[6] einzustufen, die besonders schützenswert sind. Soweit die betroffene Person, die sie betreffende besondere Kategorien von Daten offensichtlich öffentlich gemacht hat, besteht allerdings keine besondere Schutzbedürftigkeit mehr und eine Verarbeitung diese Daten kann beim Vorliegen eines Erlaubnistatbestandes gemäß Art. 6 Abs. 1 erfolgen.[7] Die Verarbeitung besondere Kategorien von Daten für Forschungszwecke ist auch ohne Einwilligung der betroffenen Personen zulässig, wenn die Verarbeitung für die Erfüllung des Forschungszweckes erforderlich ist und zusätzliche Maßnahmen zum Schutz der Rechte und Freiheiten der betroffenen Personen getroffen wurden. (u.a. § 27 Abs. 1 BDSG i. V. m. Art. 9 Abs. Abs. 2 lit. j). Grundsätzlich sollten aber besondere Kategorien personenbezogener Daten nie unverschlüsselt gespeichert und übertragen werden sowie vor unberechtigtem Zugriff geschützt werden.

---

6 Gemäß Art. 9 Abs. 1 DSGVO sind besondere Kategorien von Daten solche, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie genetische und biometrische Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten und Daten zum Sexualleben oder der sexuellen Orientierung.

7 Schulz, in: Gola Datenschutz-Grundverordnung 2018, Art. 9, Rn. 25 - 26.

5. Zusammenfassung

Die Verarbeitung personenbezogener Daten einschließlich besondere Kategorien von Daten zur Erkennung von Desinformationen für Forschungszwecke ist grundsätzlich zulässig, wenn die Rechte und Freiheiten der betroffenen Personen durch technische und organisatorische Maßnahmen (insb. Pseudonymisierung, Anonymisierung, Verschlüsselung, Zugriffskontrolle und Protokollierung) gewahrt werden. Demnach ist eine wichtige Aufgabe der forschenden Stelle, die Risiken der Verarbeitung für die Rechte der betroffenen Personen rechtzeitig zu erkennen und mit den erforderlichen Maßnahmen zu adressieren. Ist die Verarbeitung personenbezogener Daten zu einem späteren Zeitpunkt des Projekts nicht mehr erforderlich, sind die Daten zu anonymisieren bzw. zu löschen. Auch die Weiterverarbeitung der Daten für andere Forschungszwecke sollte in anonymisierter Form erfolgen. Neben den Regelungen der DSGVO hat die forschende Stelle ebenfalls das auf sie anwendbare nationale Datenschutzgesetz (z. B. das Datenschutzgesetz des Landes oder das Bundesdatenschutzgesetz) ergänzend zu beachten. Hier werden u.a. konkrete Maßnahmen zum Schutz der betroffenen Personen geregelt, die die forschende Stelle zusätzlich zu implementieren hat (siehe z. B. § 22 Abs. 2 i. V. m. § 27 BDSG).

III. Erkennung Bildmanipulation

Wenn Bilder aus unbekannten oder unseriösen Quellen auf Echtheit überprüft werden sollen, reicht eine visuelle Überprüfung durch einen Menschen heute nicht mehr aus. Ein Bild kann genau dann als nicht mehr echt angesehen werden, wenn es nachträglich bearbeitet worden ist.

1. Erkennung der Wiederverwendung

Die einfachste Art, mit Bildern Desinformationen zu betreiben, ist es, diese einfach aus ihrem Kontext heraus neu zu verwenden. Ein bekanntes Beispiel für diese Vorgehensweise sind Bilder aus Kriegsgebieten. Angeblich wird hier häufig auf Archivmaterial zurückgegriffen, da ein schnelles Beschaffen aktueller Bilder nicht oder nur unter großer Gefahr möglich ist.

Diese Wiederverwendung kann erkannt werden, wenn die Bilder über eine inverse Bildersuche in bekannten Portalen wie Google Image Search oder

TinEye gefunden werden und die Treffer in der Vergangenheit liegen. Diese Suchmaschinen sind allerdings nicht resistent gegen bewusste Verschleierung. So kann es reichen, Ausschnitte aus Bildern zu erzeugen oder diese zu spiegeln.

Sollen Bilder auch nach entsprechenden Verschleierungsschritten noch erkannt werden, sollten robuste Methoden zur Bilderkennung eingesetzt werden. Dazu gehören zum einen sogenannte „robuste Hashverfahren", im Englischen auch als image fingerprints oder perceptual hashes bezeichnet. Sie haben gemeinsam, dass sie eine effiziente Wiedererkennung von Bildern ermöglichen und resistent gegen verschiedene Veränderungen am Bild sind. Ein einfaches Beispiel ist der Blockhash aus Steinebach (2012). Hier wird ein Bild in Graustufen umgerechnet und auf 16 x 16 Pixel herunter skaliert. Aus den Pixeln wird anhand des Abstands vom Medien der Helligkeit der Pixel ein Vektor aus 256 Bit errechnet, der den robusten Hash des Bildes darstellt. Dieser Hash ist robust gegen Skalierung, leichte Bildmanipulationen und verlustbehaftete Kompression.

Ein Alternative zu den Hashverfahren sind Merkmalsvektoren wie SIFT. Sie erkennen Bilder über eine Repräsentation prägnanter Stellen wieder und sind insbesondere resistent gegen Verzerrung und Beschneiden von Bildern.

2. Erkennung von Veränderungen

Viele bestehende Algorithmen verwenden unkomprimierte Bilder für Netzwerktraining und Testing. In der Praxis ist es jedoch selten, dass unkomprimierte Bilder zugänglich sind, insbesondere für forensische Untersuchungen. In der Regel sind nur JPEG-Bilder als Original- oder manipulierte Bilder verfügbar. Um die Erkennungsleistung für JPEG-Bilder zu verbessern, haben wir ein praktischeres Szenario entwickelt, bei dem sowohl die Trainingsbilder als auch die Testbilder im JPEG-Format gespeichert werden. Wir haben die möglichen Gründe analysiert, warum das CNN-Netzwerk in [3] bei JPEG-Bildern im praktischen Szenario schlecht abschneidet. Basierend auf der Analyse haben wir ein neues Fusionsnetzwerk entwickelt, um die Erkennungsleistung bei JPEG-Bildern zu verbessern. Das Fusionsnetzwerk ist ein Verbund aus einem Inception-ResNet basierten Netzwerk und einem DCT-basierten Netzwerk, das die Stärken der beiden Netzwerke kombiniert und ihre Schwächen ausgleicht.

Das neue Fusionsnetzwerk wird zunächst im Basistest auf seine Leistungsfähigkeit bei der Identifizierung neuer Bilder mit den gleichen JPEG-

Tabelle 4.1: Ergebnisse des Basistests

	Netzwerk in [3]	**Fusionsnetzwerk**	**Δ Netzwerk in [3]-Fusion**
F1 Score	0,6958	0,9001	29,35 %
Precision	0,6916	0,9005	30,22 %
Recall	0,7001	0,8996	28,49 %

Tabelle 4.2: Ergebnisse des generalisierten Tests

	Netzwerk in [3]	**Fusionsnetzwerk**	**Δ Netzwerk in [3]-Fusion**
F1 Score	0,6973	0,8544	22,53 %
Precision	0,7022	0,8579	23,89 %
Recall	0,6925	0,8509	21,18 %

Qualitätsfaktoren wie in der Trainingsphase evaluiert. Wie in Tabelle 4.1 dargestellt, verbessert sich das Fusionsnetzwerk um etwa 30 % und die Genauigkeit erreicht 89,96 %. Weiterhin wird im generalisierten Test die Generalisierbarkeit auf neue Bilder mit unterschiedlichen Qualitäten bewertet. Die Leistung zur Erkennung von Manipulationen von Bildern mit unbekannten JPEG-Qualitätsfaktoren ist in Tabelle 4.2 dargestellt. Die Genauigkeit beträgt 85,09 % und eine Verbesserung von ca. 22 % wird erreicht. Darüber hinaus zeigt Tabelle 4.3 die Verbesserung der Erkennung für jede Art der Manipulation.

3. Erkennung von Montagen

Eine Art der digitalen Manipulation ist die Fotomontage. Eine Fotomontage kann definiert werden als die Erstellung eines Ausgangsbilds, dass mindestens Bildinhalte aus zwei verschiedenen Eingangsbildern enthält. Auch wenn die kopierten Bildinhalte prinzipiell beliebig sein können, sind für diese Aufgabenstellung eher die Bildinhalte wichtig, die sich für Desinformationen eignen. Relevante Bildinhalte sind deswegen vor allem Menschen und größere Objekte.

Der in der Praxis relevanteste Fall einer Fotomontage ist daher jener, bei dem ein relevanter Bildinhalt wie zum Beispiel eine Person in einem

Tabelle 4.3: Verbesserung der Ergebnisse für jede Art der Manipulation

	Δ F1 Score Netzwerk in [3]-Fusion	**Δ Precision Netzwerk in [3]-Fusion**	**Δ Recall Netzwerk in [3]-Fusion**
Double JPEG	125,07 %	77,05 %	168,38 %
Gaussian Blurring	8,58 %	7,99 %	9,17 %
Median Filtering	4,35 %	3,32 %	5,45 %
Resampling	17,32 %	25,59 %	8,80 %
Gaussian Noise	55,13 %	67,91 %	42,32 %

Eingangsbild in ein zweites Eingangsbild kopiert wird. So kann ein Ausgangsbild erzeugt werden, welches zum Beispiel die Person fälschlicherweise in einer Situation darstellt, die die Aussage einer ebenfalls dazu erfundenen Falschnachricht untermauert, um die Person zu diffamieren. Es ist aber auch genau das Gegenteil möglich, um zum Beispiel eine Person in einer Situation wichtig erscheinen zu lassen. Eine Fotomontage muss sich jedoch nicht nur auf das Kopieren von Fremdbildinhalten beschränken. So kann der Ersteller der Montage zusätzlich noch bestimmte Bildmanipulationen durchführen, um so das Ausgangsbild authentischer wirken zu lassen oder aber auch, um einer automatisierten Erkennung zu entgehen.

Abbildung 4.2: Ein Beispiel für eine Montage. Links findet sich die Fälschung, in der Putin auf den leeren Stuhl im Bild auf der rechten Seite einkopiert wurde. Quellen: Links Twitter, Rechts Getty Images.

Für die Erkennung von Kollagen wurde ein Ansatz gewählt, der Bilder anhand von Merkmalen vergleicht. Da Merkmale über den gesamten Bildinhalt hinweg erkannt werden können, ist es durch einen merkmalsbasierten Ansatz relativ einfach, die Ähnlichkeit ganzer Bilder oder auch einzelner Bildteile miteinander zu vergleichen.

Konzept

Die Montageerkennung selbst ist in zwei Hauptkomponenten unterteilt: Die erste Hauptkomponente ist die Initialisierung, bei der eine Bilddatenbank einmalig zu einem durchsuchbaren Index verarbeitet wird. Sie stellt also die Datenbasis dar, auf der später eine Suche durchgeführt werden kann. Im Umfeld von Desinformationen könnte dies ein Archiv von Aufnahmen sein, die bereits in der Presse verwendet wurden. Neue Bilder werden hier kontinuierlich nachgepflegt.

Zuerst wird hierzu eine Datenbank für die Bildkennungen (beispielsweise Dateinamen und Dateipfad in einem Archiv) und ihre Merkmale angelegt. Für die Generierung der Merkmale verwenden wir die Verfahren SIFT und SURF. Mit einem Detektor für diese Merkmale werden die Schlüsselpunkte für jedes Bild erstellt. Anschließend erfolgt eine Filterung, die die Anzahl der zu verwendenden Schlüsselpunkte reduziert. Anschließend werden die Deskriptoren der einzelnen Schlüsselpunkte ermittelt. Deskriptoren entstehen aus der Verarbeitung der Merkmale und stellen Vektoren dar, die die Merkmale in einer normalisierten und somit robust vergleichbaren Form speichern. Aus den Deskriptoren wird dann mit einer ausgewählten Indizierungsmethode ein Index erstellt.

Die zweite Hauptkomponente ist die Abfrage, bei der ein Eingangsbild verarbeitet und mit dem Index verglichen wird. Das Ergebnis der Verarbeitung kann dabei sowohl für die Abfrage verwendet werden, aber auch für das Einfügen des Bildes in die Datenbank aus der Indexierung, falls das Bild noch nicht bekannt ist.

Zunächst wird hierzu ein Eingabebild bereitgestellt, für welches eine Merkmalserkennung durchgeführt wird. Wie bei der Initialisierung wird dann die Filterung durchgeführt und die Deskriptoren extrahiert. Anschließend erfolgt der Abgleich der Deskriptoren mit den Deskriptoren im Index über das Indizierungsverfahren. Das Ergebnis ist eine Reihe von Übereinstimmungen, die Deskriptoren einander zuordnen. Eine Übereinstimmung besteht aus einem Deskriptor aus dem Ausgangsbild und dem ähnlichsten Deskriptor aus

einem Bild in der Datenbank. Wenn genügend Übereinstimmungen gefunden wurden, die sich auf das gleiche Bild beziehen, ist es wahrscheinlich, dass ein Objekt erkannt wird. In einem weiteren Schritt wird dies jedoch nochmals überprüft, indem die Merkmale im Detail hinsichtlich ihrer geometrischen Ähnlichkeit miteinander verglichen werden.

Optimierung der Merkmale

Eine Montage besteht aus einem oder mehreren Objekten aus verschiedenen Bildern. Die Erkennung der einzelnen Objekte erfolgt über die Merkmalserkennung. Merkmalsdetektoren reagieren in der Regel besonders auf inhomogene Oberflächen und finden Merkmale auf Bildern ohne homogene Oberflächen über das gesamte Bild. Homogene Oberflächen sind monotone Oberflächen ohne Struktur, wie z. B. ein wolkenloser Himmel oder ein niedrig aufgelöstes Bild einer Straße. Der weit verbreitete Merkmalsdetektor SIFT findet 30.000 - 40.000 Merkmale auf einem 1000 x 1000 Pixel Bild ohne homogene Flächen. Eine so hohe Anzahl von Features ist in unserem Anwendungsfall unnötig und würde dazu führen, dass der Bildindex bis zu einem Punkt wächst, der den Speicherverbrauch inakzeptabel macht.

Deshalb verwenden wir eine Filtermethode, um nur eine kleine Anzahl von Features auszuwählen und den Rest zu verwerfen. Merkmalsdetektoren bewerten die Stärke der gefundenen Merkmale. Mit einem geeigneten Filterverfahren kann sichergestellt werden, dass bei beiden Bildern nur die stärksten Merkmale ausgewählt werden und die passende Quote nicht reduziert wird.

Eine einfache Filterung der Merkmale führt aber schnell zu einem Problem im Zusammenhang mit der Montageerkennung. Wenn nur die stärksten Merkmale erhalten bleiben, bleiben oft keine oder nur noch sehr wenigen Merkmale in Teilen eines Bildes übrig. Bei einer Montage kann es nun vorkommen, dass es keine oder nicht genügende Merkmale auf einem Objekt gibt und dieses Objekt nicht mehr als kopierter Bildinhalt erkannt werden kann. Daher muss die Filterung eine Verteilung der Merkmale über das gesamte Bild gewährleisten. Durch die gleichmäßige Verteilung der Merkmale ist es sehr wahrscheinlich, dass nach der Filterung noch genügend Merkmale auf dem Objekt verbleiben. Im Idealfall reduziert dies die Anzahl der Features, ohne eine verminderte Erkennungsrate zu verursachen.

Die Anzahl der pro Bild zu verwendenden Merkmale ist eine der wichtigsten Parameter. Sie hat einen starken Einfluss auf den Abruf auf der einen Seite und auf den Speicherbedarf auf der anderen Seite. Daher muss hier ein

geeigneter Kompromiss gefunden werden. Um den Speicherverbrauch zu reduzieren, wird bei der Indizierung und Datenbankerstellung eine geringere Anzahl von Merkmalen gespeichert als bei der Suche.

In unserer Implementierung stellte sich heraus, dass 500 Merkmale bei der Indexierung und 2000 Merkmale der Suche den besten Kompromiss darstellen. Dies führt immer noch zu einer hohen Erkennungsrate, aber gleichzeitig auch zu einem akzeptableren Speicherverbrauch.

Abbildung 4.3: In der Datenbank werden Bilder und Merkmale gespeichert. Die Abfrage erzeugt aus dem Abfragebild eine Menge von Merkmalen. Diese werden in der Datenbank gesucht. Eine Übereinstimmung zeigt, dass Teile eines Bildes in der Datenbank in dem Bild der Abfrage vorkommen.

Evaluierung

Zur Evaluierung des Ansatzes wurde eine synthetische Erstellung von Kollagen implementiert, die eine ausreichend große Menge an Testmaterial erstellen konnte. Dazu wurden jeweils Objekte aus einer freien Bibliothek in ein Hintergrundbild kopiert. So entstand ein Bild mit dem Objekt, zu dem aber gleichzeitig bereits ein perfekt ausgeschnittenes Objekt bekannt war. Dieses Objekt wurde nun in ein zweites Hintergrundbild kopiert. Dieser Vorgang simuliert die Montage: In der Praxis würde das Objekt aus dem ersten Bild entnommen und in das zweite eingefügt. In der Datenbank sind

nun zwei Bilder gespeichert: Das erste Hintergrundbild mit dem Objekt und das zweite Hintergrundbild. Die Abfrage erfolgt dann mit dem zweiten Hintergrundbild, in welches das Objekt hineinkopiert wurde. Nun muss der Algorithmus in der Lage sein, das Objekt aus dem Abfragebild im ersten Bild der Datenbank zu finden und den Hintergrund des Abfragebildes als das zweiten Hintergrundbild in der Datenbank identifizieren. Um den Vorgang zu erschweren, können Bilder und Objekte verrauscht, gedreht und skaliert werden.

Ergebnisse

Die Evaluierung belegt, dass eine Erkennung von Montage mit der beschriebenen Vorgehensweise sehr zuverlässig erfolgen kann. Die folgende Tabelle zeigt einige ausgewählte Ergebnisse für die Erkennung nach verschiedenen Operationen wie Skalierung, Rotation und Rauschen, jeweils mit unterschiedlicher Stärke. Die Bilder hatten dabei eine Auflösung von 1000 mal 1000 Pixeln. Es wurde durchgehend eine Precision von mindestens 0,99 erreicht, es werden also fast ausschließlich tatsächliche Montagen vom System erkannt.

4. Erkennung von Deepfakes

Für die Erkennung von Deepfake-Angriffen wurde ein in der Fachliteratur bekannter Bildforensik-Ansatz, der sog. JPEG-Ghost-Effekt, für Video weiterentwickelt.

JPEG-Ghost-Effekt

Dieser Effekt wurde ursprünglich für digitale (Einzel-) Bilder vorgestellt (siehe Farid, 2009). Er tritt auf, wenn Bilder, Videos oder Ausschnitte davon mehrfach JPEG-komprimiert werden.

Angenommen, ein gegebenes Bild liegt als JPEG-Datei in der Qualitätsstufe Q1 vor. Letztere wird oftmals auf einer Skala von 0 (starke Kompression, entspricht schlechter Bildqualität, dafür kleiner Dateigröße) bis 100 (entspricht sehr guter Bildqualität) angeben. Von diesem Bild wird jetzt testweise durch eine weitere JPEG-Kompression bei der Qualitätsstufe Q2 eine neue Version erzeugt. Voruntersuchungen der Originalautoren haben gezeigt, dass

Tabelle 4.4: Beispiele Precision und Recall bei Manipulation von Montagen

		Precision		Recall	
		SIFT	**SURF**	**SIFT**	**SURF**
Bild Skalieren	50 %	0,9986	0,9993	0,9567	0,9913
	40 %	0,9993	0,9987	0,8947	0,9893
	30 %	1	0,9993	0,798	0,98
	20 %	0,9988	0,9993	0,546	0,9433
Rotation	10°	0,9993	0,9993	0,9367	0,974
	-10°	0,9972	0,9993	0,958	0,97
	20°	0,9993	0,9993	0,9433	0,9533
	-20°	0,9993	0,9993	0,9453	0,956
	30°	0,9993	0,9993	0,9373	0,9427
	40°	0,9993	0,9986	0,9327	0,9367
	60°	0,9979	1	0,946	0,9287
	90°	0,9951	0,9986	0,956	0,9787
	180°	0,9972	0,9993	0,9527	0,966
Rauschen	Kein	0,9986	0,9993	0,9567	0,9567
	Schwach	0,9986	0,9993	0,9527	0,9527
	Medium	0,9979	1	0,944	0,944
	Stark	0,9965	0,9993	0,9493	0,9493

sich die Farbwerte der Bildpixel dieser beiden Bildversionen am wenigsten unterscheiden, wenn Q2 = Q1 gewählt wird. Im Umkehrschluss werden sich die beiden Versionen Pixel für Pixel immer stärker unterscheiden, je „stärker“ die zweite Kompression ist (also je ausgeprägter Q2 < Q1 ist). Berechnet man also das „Differenzbild“ der beiden Versionen durch pixelweise Subtraktion der Farbwerte, wird dieses daher umso dunkler (entspricht kleinerer Differenz) je ähnlicher Q1 und Q2 sind.

Deepfake-Detektion

Für die Anwendung auf Deepfake-Videos wird dies sinngemäß auf Video-enkodierte Daten übertragen. Auch bei Video-Kompression, etwa im H.264-Codec, sind verschiedene Qualitätsstufen technisch möglich und der Ghost-Effekt kann ebenfalls beobachtet und genutzt werden. Dies wird im vorgestellten Verfahren folgendermaßen angewendet:

1. Input: Ausgangspunkt ist die zu untersuchende Videodatei. Diese kann Bildbereiche mit eingefügten Deepfake-manipulierten Gesichtern enthalten – oder auch nicht.
2. Video-Dekodierung: die Input-Videoframes (Qualität Q1 unbekannt) werden als Einzelbilder dekodiert und temporär abgespeichert.
3. Re-Enkodierung: Hieraus werden temporäre Videodateien re-enkodiert, wobei man ihre Qualitätsstufe Q2 über viele Stufen hinweg iteriert
4. Vergleich: für jede Q2-Stufe wird das Differenzbild der Input-Videoframes und des temporären Videoframes berechnet.
5. Ghost-Effekt: Bei derjenigen Q2-Stufe, bei der das Differenzbild insgesamt am dunkelsten ist, wurde demnach die Inputvideo-Qualität am besten „erraten" ($Q2 \approx Q1$). Dieses Differenzbild zeigt also den Ghost-Effekt.
6. Detektion: Falls sich im Inputvideo auch Deepfake-manipulierte Bildbereiche befunden haben, werden diese im Differenzbild als heller sichtbare Bereiche signalisiert. Dieses Graustufen-Differenzbild wird abschließend gegen einen geeigneten Schwellwert in eine reine Schwarz-Weiß-Darstellung binarisiert.
7. Gesichtserkennung: das binarisierte Differenzbild wird nur in Bildbereichen ausgewertet, die überhaupt ein Gesicht zeigen. Hierzu wird, parallel zur Schritt 2.-6., eine Gesichtserkennung durchgeführt.

Die Gesichtserkennung ist notwendig, da die Untersuchungen gezeigt haben, dass im Bildhintergrund oftmals Fehlalarme beobachtet werden. Eine genauere Analyse (z. B. bei Videos aus Nachrichtensendungen) hat gezeigt, dass diese häufig von eingeblendeten Bildern, Laufschriften oder computergenerierten Inhalten ausgelöst werden. Für die Gesichtserkennung werden externe Bibliotheken genutzt, so etwa von Bulat (2019) oder Geitgey (2019).

Es muss gesagt werden, dass die Detektion nur erfolgreich ist, wenn die eingefügten Bildinhalte bei einer von Q1 abweichenden Qualitätsstufe enkodiert worden waren, also einem anderen Kompressions-„Lebenszyklus" unterworfen waren als der unveränderte Hintergrund.

Beispiel

Die Effektivität des Verfahrens kann man an den folgenden Beispielbildern erkennen: das erste Bildpaar zeigt ein Einzelbild des Originalvideos und des Deepfake-manipulierten Videos.

Abbildung 4.4: links: originaler Video-Frame; rechts: Deepfake-Manipulation und Gesichtserkennung (rot)

Im folgenden Bildpaar ist hierfür das Detektionsergebnis zu sehen. Man kann erkennen, dass die als Fälschung signalisierten Pixel im tatsächlich manipulierten Bild (rechts) viel dichter liegen als im unverfälschten Original-Frame (links). Durch geeignete Wahl eines Schwellwerts für diese Dichte wird eine eindeutige Klassifikation möglich.

Abbildung 4.5: links: Detektionsergebnis für original Video-Frame, rechts: für Deepfake-Mainpulation

IV. Erkennung von Desinformationen in Texten

Um eine linguistische Analyse vornehmen zu können, um darauf aufbauend maschinelle Lernverfahren zu trainieren, bedarf es eines Gegenkorpus (zu den Fake-News-Referenztexten aus Kap. 2.B) mit inhaltlich wahren Nachrichten. Diese wurden von den Online-Auftritten deutscher Zeitungen wie Süddeutsche Zeitung und Frankfurter Allgemeinen Zeitung bezogen (gecrawlt). Es wurden nur Artikel gecrawlt, die im gleichen Zeitraum wie die Referenztexte publiziert wurden. Zudem wurde darauf geachtet, dass die Artikel thematisch

mit dem Inhalt der verifizierten Fake News zusammenhängen (z. B. „Migration", „Innere Sicherheit", „Europapolitik", „US Präsidentschaftswahl 2016", etc.). Potthast et al. (2018) hatten in einer Studie gezeigt, dass keiner der von ihnen untersuchten Mainstream-Texte komplett falsch war. 97,5 % der Nachrichten aus Mainstream-Texten basieren auf seriöser und wahrhaftiger Berichterstattung. Nur 8 von 826 Texten (ca. 1 %) hatten eine Mischung aus korrekten und falschen Angaben. Diese Abweichung wurde akzeptiert und sich auf die Wahrhaftigkeit der Artikel verlassen.

Die Untersuchungen haben einige Forschungsergebnisse, welche auf englischen Daten vorgenommen wurden, bestätigt, andere konnten wiederlegt werden. Fake-News-Titel sind durchschnittlich länger als die Titel von True News. Untersuchungen haben gezeigt, dass vor allem Social-Networks-Nutzende (z. B. auf Facebook) dazu tendieren nur den Titel statt den gesamten Artikel zu lesen (Horne & Adali, 2017). Aus diesem Grund werden Fake News-Artikel, welche eine bestimmte politische Agenda vorantreiben, so verfasst, dass so viel Inhalt wie möglich im Titel zusammengefasst wird. Die Verwendung von Großbuchstaben kann hauptsächlich in Titeln von Clickbaiting-Artikeln beobachtet werden, welche versuchen, auf diese Weise Aufmerksamkeit des Lesers auf sich zu ziehen.

Potthast et al. (2018) haben gezeigt, dass der Textkörper von Fake News in der Regel kürzer ist als der von Mainstream-News. Unsere Forschungsergebnisse mit deutschen Daten haben ebenfalls gezeigt, dass Fake-News-Texte etwas kürzer verfasst sind, jedoch nicht signifikant. Es konnte nicht bestätigt werden, dass True News komplexere und längere Wörter verwenden (Rashkin et al., 2017).

Zudem wurde die Verteilung von Falschaussagen in Fake News analysiert. Wenn der Artikel eine Falschaussage enthält, wird diese in 45 % der Fälle bereits im Titel offeriert, in 39 % der Fälle im Textkörper. Weitere Fake-Statements können im Text verortet werden. In nur etwa 10 % der Fälle tauchen Falschaussagen im Vorspann auf oder können in Bildern verortet werden (5,3 %).

Rashkin et al. (2017) und Horne und Adali (2017) haben gezeigt, dass Fake News in ihren Artikeln häufiger Personalpronomen verwenden, um den Leser persönlicher zu adressieren. Im Projekt wurden die relativen Häufigkeiten von Personalpronomen in allen Nachrichtentexten untersucht. Fake News verwenden nicht signifikant mehr Personalpronomen. „Ich" wird beispielsweise häufiger in Nachrichten von Qualitätsmedien verwendet. Dies könnte aber damit zusammenhängen, dass diese Nachrichten häufiger Direktzitate

verwenden. Das Personalpronomen „wir“ kommt in beiden Referenzdaten gleich häufig vor.

1. Datenanalyse mit Machine Learning und Natural Language Processing Verfahren

Um die mithilfe des Crawling-Frameworks erfassten Textdaten verarbeiten und analysieren zu können, wurde im Rahmen des Projekts ein Analyse-Modul entwickelt, mit dessen Hilfe Texte strukturiert, repräsentiert und anschließend analysiert werden können.

Datenaufbereitung (Datenbereinigung und Datenstrukturierung)

Die gecrawlten Textdaten müssen vor ihrer Analyse entsprechend aufbereitet werden. Dies schließt im Wesentlichen eine Säuberung ein, die nicht auf der Metadatenebene, sondern auf der Inhaltsebene durchgeführt wird. Im Folgenden werden die einzelnen Schritte der Säuberung erläutert.

Da im Rahmen des Projekts nur deutschsprachige Texte untersucht werden sollten, galt es bezüglich der gecrawlten Datenströme nur relevante Texte zu extrahieren. Alle anderen Textdaten, deren Sprache nicht Deutsch war, wurden für die weitere Verarbeitung ausgeschlossen. Als Spracherkenner verwendeten wir eine eigene Implementierung, die auf Funktionswörter und Zeichen-N-Gramme basiert und eine nahezu perfekte Erkennung (> 99 %) bzgl. 50 Sprachen aufweist.

Nach der Extraktion deutschsprachiger Texte aus den Datenströmen galt es im nächsten Schritt die Texte entsprechend zu säubern. Hierfür verwendeten wir eine Pipeline, die die Texte sequentiell nach vorgegebenen Regeln säubert. Konkret wurden die Texte von Grundrauschen wie Hyperlinks, Überschriften, Bildsignaturen und Ähnlichen bereinigt. Des Weiteren wurden Datumsangaben, Währungen und anderweitige numerische Textstellen durch einheitliche Dummy-Tokens normalisiert (z. B. 21.5.2018 → <DATUM>, 20:12 → <UHRZEIT>). Dies hat den Effekt, dass Machine Learning Verfahren (kurz ML-Verfahren) von spezifischen Zahlen abstrahieren sollen, um einer Überanpassung (Overfitting) entgegenzuwirken, gleichzeitig aber die abstrakten Angaben nutzen können, um syntaktische Muster zu lernen, die hinsichtlich einer Unterscheidung zwischen Fake-/Nicht-Fake-Texten hilfreich sein können.

Nach Bereinigung der Textdaten galt es, diese mithilfe von Natural Language Processing (kurz NLP) Verfahren zu strukturieren, um bezüglich der Analyse auf die unterschiedlichsten Merkmale zuzugreifen. Zu solchen Merkmalen gehören z. B. Wortklassen (Funktionswörter/ Inhaltswörter), Wortarten (Nomen, Adjektive, Verben, Adverbien, etc.), Entitäten, semantische Relationen, Zitate, Redewendungen, etc. Es entstand hierfür ein gesondertes Framework, mit dessen Hilfe eine effiziente MerkmalEDsextraktion durchgeführt werden konnte. Als zugrundeliegende NLP-Werkzeuge dienten hierbei unter anderem Tokenizer, Part-of-Speech-Tagger, Chunker sowie Named Entitey Recognizer.

Datenrepräsentation

Ausgehend von dem Strukturierungsframework entstand ein weiteres Werkzeug mit dessen Hilfe Texte geeignet repräsentiert werden konnten, um diese seitens von ML-Verfahren untersuchen zu können. Hierfür entwickelten wir Repräsentationen basierend auf Bag-of-Features, Embeddings sowie Language Models. Die Repräsentationen können hierbei als Modelle aufgefasst werden, die entsprechende Texteinheiten numerische Werte zuweisen (Beispiel: absolute/relative Häufigkeiten von bestimmten Wörtern oder die Auftretenswahrscheinlichkeit eines Folgewortes in einer Sequenz von Wörtern).

Klassifikationsverfahren

Die zentrale Komponente des Analyse-Moduls sind sogenannte Klassifikatoren, mit deren Hilfe gegebene Texte automatisiert hinsichtlich vorgegebener Klassen eingeordnet werden können. Die Klassen, die im Vordergrund stehen sind „Fake“ und „Nicht-Fake“. Im Rahmen des Projekts entstanden mehrere Klassifikatoren, von denen ein Verfahren (genannt OCCAV) auf einer angesehenen internationalen Konferenz für Information Retrieval vorgestellt wurde (ECIR 2018). OCCAV basiert dabei auf einer Language Model Repräsentation und hat den markanten Vorteil, dass es ohne Training auskommt und Textdaten daher ohne Vorwissen direkt klassifizieren kann. Dies ist insbesondere in solchen Szenarien wichtig, in denen entweder keine Trainingsdaten existieren oder vorliegende Daten nicht geeignet sind. Vereinfacht ausgedrückt erlernt OCCAV ein Language Model aus einer Menge gegebener

Texte, die eine Klasse X repräsentieren, und versucht dieses Modell in einem unbekannten Text Y wiederzufinden. Wenn dies erfolgreich gelingt, wird angenommen, dass Y zur Klasse X gehört, ansonsten handelt es sich bei Y um eine andere Klasse. Übertragen auf das Projekt stellt X die Klasse der Fake-Texte, sodass wenn Y dieser Klasse zugewiesen wird, es sich ebenfalls um einen Fake-Text handelt, andernfalls um einen Nicht-Fake-Text.

Evaluierung

In einer Reihe von Experimenten testeten wir die Anwendung unserer Klassifikatoren zum Zwecke der Erkennung von Fake-News bzw. zur Diskriminierung zwischen Fake und Nicht-Fake.

Experiment: Unterscheidung von Fake News vs. True News auf Basis syntaktischer Strukturen

In diesem Experiment stellten wir einen Korpus zusammen, der 100 Klassifikationsfälle umfasste. Diese unterteilten sich in 50 Fälle, bei denen X mit Y übereinstimmt (unbekannte und bekannte Texte gehören der Klasse „Fake" an), und weitere 50 Fälle, in denen X mit Y nicht übereinstimmt (bekannte Texte = „Fake", unbekannter Text = „Nicht-Fake"). Als Klassifikator wählten wir hierbei OCCAV. Als Vorverarbeitungsschritt maskierten wir zunächst themenbehaftete Wörter, sodass nur noch syntaktische Strukturen wie Funktionswortphrasen inklusive Interpunktionszeichen in den Texten verblieben.

Beispiel

„Die fast regelmäßg gemeldeten Pannen der Bundeswehr lassen eigntlich aufhorchen, doch es scheint niemanden der Abgeordneten aus dem gleichgeschalteten Bundesparlament ernsthaft zu interessieren."

*„Die * * * * der * * * *, * es * * der * aus dem * * * zu *."*

Die Fragestellung, der somit nachgegangen werden soll, ist, ob eine Unterscheidung von Fake News gegenüber True News alleine auf Basis syntaktischer Strukturen möglich ist. Mit anderen Worten bedeutet dies, dass der Klassifikator sich nicht auf themenbehaftete Inhalte fokussieren soll. Als

Baselines wählten wir drei One-Class-Verfahren (PCA, SOS und LOCI) aus einem existierenden Framework (PyOD). Alle drei erzielten hinsichtlich des gegebenen Korpus ein zufälliges Klassifikationsergebnis (50 %). OCCAV erzielte hingegen eine Erkennungsgenauigkeit von 69 %. Dies zeigt, dass es möglich ist, eine Unterscheidung unabhängig von dem eigentlichen Inhalt der Nachrichtentexte zu erzielen, auch wenn das Ergebnis sicherlich verbesserungsfähig ist.

V. Malicious-Bot-Erkennung

Der Einsatz von Botnetzen ist ein Mechanismus, der auch im Kontext von Desinformationen zunehmend relevant ist. Wir beschreiben hier erste Ergebnisse und Ansätze einer automatisierten Bot-Erkennung.

1. Beschreibung des Datensatzes

Der bereitgestellte Trainingsdatensatz der Bots- und Gender-Profiling-Aufgabe auf PAN 2019 (Rangel et. al, 2019) besteht aus 4.120 englischen und 3.000 spanischen Twitter-Accounts. Jede dieser XML-Dateien enthält 100 Tweets pro Autor. Jeder Tweet wird in einem „Dokument“ XML-Tag gespeichert.

Jeder Autor wurde mit einer alphanumerischen Autoren-ID kodiert. Der englischsprachige Ordner enthält 2.060 Bot-Texte, 1.030 weibliche und die gleiche Anzahl männlicher Texte. Der spanische Ordner ist kleiner als der englische und umfasst 1.500 Bot-Texte und 750 Texte pro Geschlecht.

Um eine Überanpassung beim Training eines Klassifikators zu vermeiden, werden die Daten in ein Trainings- (70 %) und Test- (30 %)-Set aufgeteilt - wie von den PAN-Organisatoren empfohlen.

Bei Betrachtung des binären Klassifizierungsproblems „Bot vs. Human“ ist der Datensatz ausgeglichen. Wird sie jedoch zu einem Dreiklassenproblem umformuliert, dominiert die „Bot“-Klasse über die beiden Geschlechterklassen „männlich“ und „weiblich“. Unsymmetrische Daten beziehen sich auf eine ungleiche Verteilung von Klasseninstanzen. Dieses Ungleichgewicht kann durch den Einsatz der „Undersampling“-Technik weitgehend reduziert werden. Durch die zufällige Entfernung von Texten aus der Mehrheitsklasse ermöglicht diese einfache Methode die Erstellung ausgewogener Datensätze, die theoretisch zu eine Klassifikation führen, die nicht auf eine bestimmte

Klasse fokussiert ist. Durch das Undersampling der Bot-Klasse sind wir das Risiko eingegangen, wichtige Instanzen auszulassen, die wichtige Unterschiede zwischen den drei Klassen aufweisen können. Dadurch wurde die Anzahl der englischen Bot-Texte von 2.060 auf 1.030 reduziert, je nach Größe der Autoren und Autorinnen pro Klasse. Die spanischen Bot-Texte wurden von 1.500 auf 750 Instanzen reduziert. Zusätzlich haben wir den Trainingsdatensatz in drei kleinere Mengen aufgeteilt. 50 % der Daten wurden für das Training, 25 % für die Validierung und 25 % für die Testung der SVM verwendet.

2. Methodik

Im Folgenden wird für jede Sprache der gleiche Ansatz angewendet. Zuerst werden die Twitter-Daten vorverarbeitet, um Textbesonderheiten wie Hashtags, URLs und Benutzererwähnungen zu behandeln. Anschließend werden Wort-Unigramme und Bigramme sowie Zeichen-N-Gramme im Bereich von 3 bis 5 als Merkmale extrahiert, die als Input für das Training einer Support Vector Machine (SVM) dienen.

Vorverarbeitung

Die Vorverarbeitungspipeline ist für beide Sprachen (Englisch und Spanisch) nahezu gleich. Die folgenden Schritte werden durchgeführt, um die Tweets zu reinigen und zu strukturieren:

1. Konkatenierung aller 100 Tweets pro Autor zu einer langen Zeichenkette.
2. Kleinschreibung aller Zeichen.
3. Entfernen von Leerzeichen.
4. Ersetzen von URLs durch den Platzhalter <URL>.
5. Löschen von irrelevanten Zeichen, z. B. „+,*,/,/,“.
6. Ersetzen aller Hashtags und angehängter Token durch den Platzhalter <HashTag>. 7. Ersetzen von @-Mentions (z. B. @username) durch den Platzhalter <UsernameMention>.
7. Sequenzen mit gleichen Zeichen und einer Länge von mehr als drei werden entfernt.
8. Entfernen von Wörtern mit weniger als drei Zeichen.

9. Entfernen von Stoppwörtern mit Hilfe der NLTK (Natural Language Toolkit) Bibliothek.
10. Um die Wörter zu tokenisieren, haben wir den TwitterTokenizer aus der NLTK-Bibliothek verwendet.

Merkmale

Da die beiden Sprachen unterschiedliche Datensätze haben, wurden zwei separate Klassifikationsmodelle für jede Sprache trainiert. Wir haben verschiedene Feature-Sets getestet und mit Hyperparameter-Tuning experimentiert, manuell und mit der Grid-Suchfunktion von scikit-learn. Die Hyperparameter wurden für jedes Sprachmodell separat abgestimmt. Verschiedene Experimente werden in Abschnitt 6 diskutiert.

Nach der Vorverarbeitung wurde eine Wortfrequenzanalyse auf beiden Datensätzen durchgeführt. Wir haben die Trainings-, Validierungs- und Testsets zusammengeführt. Die drei am häufigsten verwendeten Token von Bots sind:

a. URLs (Token <URL>)
b. Hash-Tags (Token < HashTag>)
c. und @-Mentions (Token < UsernameMention >)

Während Bots eine höhere Neigung haben, URLs zu teilen, neigen Menschen dazu, sich in erster Linie auf andere Nutzende (oder Konten) zu beziehen, indem sie @-Mentions verwenden (markiert als <UsernameMention>-Token). Neben dem Verweis auf andere Nutzende verwenden Menschen am zweithäufigsten URLs. Der dritthäufigste Token, den Menschen auf Twitter verwenden, ist die Verwendung von Hashtags. Diese Analyse zeigt, dass diesen Token besondere Aufmerksamkeit geschenkt werden sollte, wenn Twitter-Texte vorverarbeitet werden.

Je nach Häufigkeitsverteilung wurden die 10.000 am häufigsten verwendeten Token im Trainingsset in einem Dictionary gespeichert. Beim Aufbau des Vokabulars wurden Begriffe mit einer Dokumentenfrequenz von weniger als 2 ignoriert.

Es wurden TF-IDF zum Vektorisieren verwendet, um eine Vektor-Pipeline für jede Sprache aufzubauen. Die folgenden N-Gramme für beide Sprachen wurden verwendet:

a. Wortunigramme und Bigramme

b. Zeichen-N-Gramme im Bereich von 3 bis 5

Die Art und Weise, wie die Wort- und Zeichenauswahl durchgeführt wurde, ist von Daneshvar und Inkpen (2018) inspiriert. Die Autoren präsentierten ihren Gender Identification-Ansatz für Twitter-Texte bei der PAN Challenge im Jahr 2018, bei der ihr Modell auf dem zweiten Platz landete.

3. Algorithmus des maschinellen Lernens

Um einen Klassifikator zu trainieren, verwendeten wir eine lineare SVM mit verschiedenen Wort- und Zeichen-N-Grammen als Features. Da wir die Aufgabe als ein Multiklassen-Problem betrachten, wurde die Entscheidungsfunktion OVR („One-vs.-Rest") verwendet. OVR kombiniert mehrere binäre SVMs zur Lösung der Klassifikationsaufgabe mit dem Training einer Vielzahl von Klassen. Die drei zu trainierenden Klassen sind: „Bot", „Männlich" und „Weiblich". Mit OVR klassifiziert jede SVM eine Klasse gegen alle anderen Klassen.

Um eine Überanpassung beim Experimentieren mit dem Trainingsset zu vermeiden, haben wir die von den Veranstaltern zur Verfügung gestellten Daten in drei Teile gegliedert. Für das Training haben wir 50 % der Daten verwendet. Die andere Hälfte des Datensatzes wurde zu gleichen Teilen als Validierungs- und Testsatz aufgeteilt (jeweils 25 % der Textdaten). Während der Experimente sah das Modell den Testsatz nicht. Die Parametereinstellung wurde am Validierungsdatensatz durchgeführt. Schließlich wurde jedes Modell auf dem offiziellen PAN 2019 Testset für den Author-Profiling-Task auf der TIRA-Plattform getestet.

VI. Ergebnisse

Die folgende Tabelle zeigt die Ergebnisse des Verfahrens, die mit dem vorläufigen Trainingsset erzielt wurden, sowie die Genauigkeitswerte mit dem offiziellen Testset. Die Genauigkeitswerte wurden für jede Sprache einzeln berechnet. Zuerst wurde die Genauigkeit bei der Identifizierung von Bot und Mensch berechnet. Dann, im Falle eines Menschen, wurde die Genauigkeit der Vorhersage ob Mann oder Frau berechnet. Jedes Modell wurde auf 50 % der Testdaten trainiert. Die Hyperparameter wurden auf dem 25 %igen Vali-

Tabelle 4.5: Genauigkeitswerte für Bot- und Geschlechtererkennung am „Early Bird“ und am offiziellen PAN 2019 Testdatensatz

	„Early Bird“		Testdatensatz	
Sprache	DE	ES	DE	ES
Bot vs. Mensch	0,97	0,97	0,92	0,91
Männlich vs. weiblich	0,94	0,93	0,82	0,78

dierungssplit angepasst. Schließlich wurde das eingereichte Modell auf dem offiziellen PAN 2019 Testset auf der TIRA-Plattform getestet.

VII. Weitere getestete Methoden und Merkmale

Neben den bereits beschriebenen Schritten der Vorverarbeitung und Merkmalsauswahl wurden auch andere Merkmale und Datenstrukturtechniken untersucht. Neben der vorgestellten SVM mit einem linearem Kernel wurden auch andere Klassifikatoren getestet, nämlich CNN und den Random Forest Classifier. In den Experimenten konnten diese beiden Klassifikatoren in Bezug auf die Leistung nicht mit der linearen SVM mithalten.

In den Experimenten wurden Twitter-Daten wie folgt bereinigt: Entfernung aller URLs, Hashtags, Retweets (RT) und @-Mentions. Experimente haben gezeigt, dass diese Features für die Bot-Erkennung von Twitter-Daten unerlässlich sind. Um die Token zu vektorisieren, wurde zunächst mit gesamten und relativen Wortfrequenzen sowie mit der Konvertierung der Tokens in tf-idf gearbeitet. Die Vektorlänge lag zwischen 1.000 und 10.000 der häufigsten vorkommenden Token. Die Experimente zeigten, dass die Genauigkeit abnahm, wenn die Hyperparameter mit der Rastersuchfunktion von scikit-learn angepasst wurden. Die Tabelle zeigt die Ergebnisse der Experimente mit dem Testdatensatz „Early Bird“.

C. Diskussion und Zusammenfassung

Die große Menge von Meldungen, die potentiell Desinformationen enthalten können, macht eine Unterstützung von menschlichen Beobachtern durch technische Maßnahmen notwendig. Diese Maßnahmen können heute noch nicht selbständig eine Aussage über Desinformationen machen und somit als

Tabelle 4.6: Genauigkeitswerte für Bot- und Geschlechtererkennungsexperimente auf dem Datensatz des PAN 2019 „Early Bird“ Testdatensatz.

	Bot vs. Human		Male vs. Female	
Sprache	DE	ES	DE	ES
Token Gesamthäufigkeit	0,91	0,83	0,65	0,64
Token Relative Frequenz (Grid Search Tuning)	0,73	0,83	0,53	0,64
Token TF-IDF Vektorisierung	0,92	0,78	0,81	0,61

autonomer Filter Kommunikationskanäle überwachen. Sie können aber einen Redakteur, der den Wahrheitsgehalt einer Meldung betrachtet, unterstützen.

Verschiedene Verfahren sind heute bereits so ausgereift, dass sie problemlos in der Praxis eingesetzt werden können. Das Wiedererkennen von Bildern und das Erkennen der Bestandteile von Bildmontagen weist nur noch Fehlerraten im Bereich von unter einem Promille und wenigen Prozent auf. Eine Nutzung erfordert hier allerdings den Aufbau entsprechender Referenzdatenbanken. Nur wenn die Bilder zuerst in einer Datenbank gespeichert sind, können sie auch wiedererkannt werden. Unterstützen können hier automatisierte Crawler, die selbständig Bilder finden und in die Datenbank einspeisen.

Andere Verfahren wie die Manipulationserkennung von Bildern und das Erkennen von Deepfake-Videos sind auf einem guten Weg, hier ist die Erkennung allerdings noch nicht so weit fortgeschritten, dass sie einfach einzusetzen sind. Sie können Hinweise geben und in geeigneten Fällen auch sehr präzise Bewertungen durchführen, sind aber noch nicht in der Lage, in allen Fällen eine Manipulation zu erkennen und neigen andererseits auch dazu, Fehlalarme auszulösen. Hier muss also der Anwender die Ergebnisse interpretieren und genug Fachkenntnis besitzen, eine abschließende Entscheidung zu treffen. Dies gilt ebenso für die Erkennung von Texten; sowohl linguistische Methoden als auch Ansätze aus dem Maschinellen Lernen zeigen, dass eine Erkennung von Desinformation und ähnlichen Inhalten mit hohen Trefferquoten möglich ist. Trotzdem muss bei Fehlerraten von 30 Prozent und mehr das abschließende Urteil von einem Anwender erfolgen.

Um hier in der Zukunft bessere Ergebnisse zu erzielen, ist das Schaffen einer besseren Trainingsgrundlage von großer Bedeutung. Maschinelles Lernen wird im Kontext von Desinformation erst dann wirklich erfolgreich sein können, wenn große Menge von Texten und auch anderen Medien, die als

Desinformation erkannt wurden, in Datenbanken abgelegt und entsprechend kommentiert sind. Auf diesen Daten können dann zukünftige Netze trainiert werden.

Die Nutzung erster Ergebnisse durch im Journalismus Tätige, also Fachanwender, kann in naher Zukunft erfolgen. Weiter in die Zukunft blickend ist aber auch eine Verbesserung der Verfahren bis zu einem Grad der Automatisierung denkbar, der die Methoden auch für den einzelnen Bürger verfügbar macht. Eine Integration in Browser würde dann Hinweise geben, dass beispielsweise eine Text einen Stil aufweist, in dem in der Vergangenheit schon zahlreiche belegte Desinformationen verfasst wurden, oder dass ein betrachtetes Bild Spuren von Manipulationen aufweist.

Ein automatisiertes Blockieren und Löschen von Inhalten durch entsprechende Verfahren hingegen wird immer problematisch sein, da hier Maschinen eine Aufgabe erhalten, die in der Praxis immer Fehlerraten aufweisen wird. Denkbar ist aber, dass im Falle der Verfügbarkeit entsprechender Methoden und der damit einhergehenden Erleichterung bei der Prüfung von Meldungen der Druck auch auf die Verbreiter von Nachrichten steigt, hier eine verantwortungsvolle Prüfung durchzuführen.

Literaturverzeichnis zu Kapitel 4

Afchar, D. und Nozick, V. und Yamagishi, J. & Echizen, I (2018). MesoNet: a Compact Facial Video Forgery Detection Network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE.

Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. In 2012 IEEE Symposium on Security and Privacy (pp. 461-475). IEEE.

Ahmed, H. (2017). Detecting opinion spam and fake news using n-gram analysis and semantic similarity, Ph.D. thesis, University of Victoria. https://dspace.library.uvic.ca//handle/1828/8796.

Allcott, H., Gentzkow, M. (2017), Social Media and Fake news in the 2016 Election. In: Journal of Economic Perspectives. 31(2), 211-236.

Bacciu, A., La Morgia, M., Mei, A., Nemmi, E. N., Neri, V., & Stefa, J. (2019). Bot and Gender Detection of Twitter Accounts Using Distortion and LSA. CLEF (Working Notes)

Banko, M., Etzioni, O., Soderland, S., Weld, D. S. (2008), Open information extraction from the web. IJCAI, Vol. 7, (pp. 2670-2676).

Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In European conference on computer vision (pp. 404-417). Springer, Berlin, Heidelberg.

Bayram, S., Sencar, H. T., & Memon, N. (2008, September). A survey of copy-move forgery detection techniques. In IEEE Western New York Image Processing Workshop (pp. 538-542). IEEE.

Belhassen Bayar and Matthew C. Stamm, A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer, In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, (pp. 5–10), 2016.

Belhassen Bayar and Matthew C. Stamm, Design Principles of Convolutional Neural Networks for Multimedia Forensics, Electronic Imaging, Media Watermarking, Security, and Forensics 2017, pp. 77-86(10), 2017.

Bianchi, T. und Piva, A. (2012). Image Forgery Localization via Block-Grained Analysis of JPEG Artifacts. IEEE Transactions on Information Forensics and Security. vol. 7, issue 3, S. 1003 ff. IEEE

Birajdar, G. K., & Mankar, V. H. (2013). Digital image forgery detection using passive techniques: A survey. Digital investigation, 10(3), 226-245.

Bulat, A. (2019) 2D and 3D Face alignment library build using pytorch. In: Github repository '1adrianb/face-alignment'. https://github.com/1adrianb/face-alignment

Castillo, C., Mendoza, M., Poblete, B. (2011). Information credibility on twitter. Proceedings of the 20th international conference on world wide web. ACM675–684.

Cavoukian, A. (2009). Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada, 5.

Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. Technischer Report. In: arXiv.org, Cornell University, Cornell University, Ithaca, NY, USA

Daneshvar, S., Inkpen, D.: Gender identification in twitter using n-grams and lsa: Notebook for pan at clef 2018. In: CLEF (2018)

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., Menczer, F. (2016). Botornot: A system to evaluate social bots. Proceedings of the 25th international conference companion on world wide web. International World Wide Web Conferences Steering Committee273–274.

De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006, May). Generating typed dependency parses from phrase structure parses. In Lrec (Vol. 6, pp. 449-454).

deepfakes (Github user) (2019). Non official project based on original /r/Deepfakes [Reddit] thread. In: Github repository 'deepfake/faceswap'. https://github.com/deepfakes

Deepfakes web beta (2019). Create your own Deepfakes online. https://deepfakesweb.com/

Dickerson, J. P., Kagan, V., Subrahmanian, V. (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? Advances in social networks analysis and mining (asonam), 2014 IEEE/ACM international conference on. IEEE620–627.

Farid, H. (2009). Exposing Digital Forgeries From JPEG Ghosts. IEEE Transactions on Information Forensics and Security, vol. 4, issue 1, S. 154 ff., IEEE

Geitgey, A. (2019) The world's simplest facial recognition api for Python and the command line. In: Github repository 'ageitgey/face_recognition'. https://github.com/ageitgey/face_recognition

Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., Crowcroft, J.: Of bots and humans (on twitter). In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. pp. 349–354. ACM (2017)

Gilani, Z., Wang, L., Crowcroft, J., Almeida, M., Farahbakhsh, R.: Stweeler: A framework for twitter bot analysis. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 37–38. International World Wide Web Conferences Steering Committee (2016)

Hany Farid, „Image forgery detection", IEEE Signal Processing Magazine, vol. 26, issue 2, pp. 16-25, 2009.

Horne, B. D., Adali, S. (2017): This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In: International AAAI Conference on Web and Social Media, 759-766.Jin, Z., Cao, J., Zhang, Y., Luo, J. (2016). News verification by exploiting conflicting social viewpoints in microblogs. AAAI2972–2978.

Jason Bunk, et al, „Detection and Localization of Image Forgeries using Resampling Features and Deep Learning", CVPR Worshop on Media Forensics, July 2017.

Karataş, Arzum & Şahin, Serap. (2017). A Review on Social Bot Detection Techniques and Research Directions. ISCTurkey 10th International Information Security and Cryptology Conference, At Ankara, Turkey

Kowalski, M. (2018). 3D face swapping implemented in Python. In: Github repository 'MarekKowalski/FaceSwap'. https://github.com/MarekKowalski/FaceSwap

Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y. (2013). Prominent features of rumor propagation in online social media. Data mining (icdm), 2013 ieee 13th international conference on. IEEE1103–1108.

Li, W. und Yuan, Y. und Yu, N. (2009). Passive detection of doctored JPEG image via block artifact grid extraction. Signal Processing. vol. 89, issue 9, S. 1821 ff. Elsevier

Li, Y. und Chang, M.-C. und Lyu, S. (2018). In Ictu Oculi: Exposing AI Created Fake Videos by Detecting. Eye Blinking. Technischer Report. University at Albany, State University of New York, NY, USA.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

Luo, W., Huang, J., & Qiu, G. (2010). JPEG error analysis and its applications to digital image forensics. IEEE Transactions on Information Forensics and Security, 5(3), 480-491.

Niu, X. M., & Jiao, Y. H. (2008). An overview of perceptual hashing. Acta Electronica Sinica, 36(7), 1405-1411.

Pan, J. et al. (2018): Content based fake news detection using knowledge graphs. In International Semantic Web Conference (pp. 669-683). Springer, Cham.

Pawel Korus, „Digital image integrity – a survey of protection and verification techniques", Digital Signal Processing 71, pp. 1 –26, 2017.

Potthast, M., Hagen, M., Stein, B. (2016): Author Obfuscation: Attacking the State of the Art in Authorship Verification. In: CLEF (Working Notes), 716-749.

Rangel, F., Rosso, P.: Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling. In: Cappellato, L., Ferro, N., Losada, D., Müller, H. (eds.) CLEF 2019 Labs and Workshops, Notebook Papers. CEUR-WS.org.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., Choi, Y. (2017): Truth of Varying Shades: Analyzing Language in Fake news and Political Fact-Checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, (pp. 2931-2937).

Raskin, V. (1987). Linguistics and natural language processing. Machine translation: Theoretical and methodological issues.. Cambridge University Press, Cambridge (pp. 42–58).

Rössler, A. und Cozzolino, D. und Verdoliva, L., Riess, C., Thies, J., & Nießner, M (2018). FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces. Technischer Report. In: arXiv.org, Cornell University, Ithaca, NY, USA

Rubin, V., Chen, Y., Conroy N. J. (2015): Deception detection for news: three types of fakes. In: Proceedings of the 78th ASIS & T Annual Meeting: Information Science with Impact: Research in and for the Community (ASIST '15). American Society for Information Science, Silver Springs, MD, USA, Article 83.

shaoanlu, clarle (Github users). (2019). A denoising autoencoder + adversarial losses and attention mechanisms for face swapping. In: Github repository 'shaoanlu/faceswap-GAN'. https://github.com/shaoanlu/faceswap-GAN

Shin, J., & Ruland, C. (2013, October). A survey of image hashing technique for data authentication in WMSNs. In 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 253-258). IEEE.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.

Steinebach, M., Liu, H., & Yannikos, Y. (2012, February). Forbild: Efficient robust image hashing. In Media Watermarking, Security, and Forensics 2012 (Vol. 8303, p. 83030O). International Society for Optics and Photonics.

Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv:170407506.

Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A., (2017). Online human-bot interactions: Detection, estimation, and characterization. In: Eleventh international AAAI conference on web and social media.

Yang, J., Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in twitter. ICWSM, 10(2010), 355–358.

Zhang, X., Ghorbani, A. A. (2019). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management.

Kapitel 5: Desinformation aus der Perspektive des Rechts

Autoren:

Lena Isabell Löber
Prof. Dr. Alexander Roßnagel

Desinformation ist gesetzlich nicht definiert. Aus rechtlicher Perspektive ist sie als bewusst unwahre Tatsachenbehauptung zu qualifizieren. Sie kann auch mit Werturteilen verbunden sein. Neben das objektiv zu bestimmende Kriterium der Unwahrheit tritt eine subjektive Komponente: Die Äußerung der Unwahrheit erfolgt wider besseres Wissen und zumeist mit Täuschungs- oder Manipulationsabsicht. Die Verbreitung der Desinformation erfolgt in einem Kommunikationszusammenhang, auf den sie einwirkt oder einzuwirken versucht.[1] Die Weiterverbreitung der Desinformation kann durch Personen geschehen, denen die Unwahrheit nicht bewusst ist.

Ein generelles Gesetz, das die Herstellung und Verbreitung von Desinformationen verbietet oder unter Strafe stellt, gibt es in Deutschland nicht.[2] Ein solches Verbot besteht nur in bestimmten Kommunikationszusammenhängen, wie z. B. vor Gerichten und parlamentarischen Untersuchungsausschüssen[3] oder zum Schutz des Wettbewerbs, der Verbraucher oder von Vertragspartnern. Auch untersagt das Strafrecht gewisse Desinformationen, die für die Allgemeinheit und Einzelne besonders schädlich sind.

Die folgenden Ausführungen untersuchen die bestehenden rechtlichen Regelungen und die rechtspolitischen Regulierungsansätze, die diskutiert werden, um digitale Desinformation und ihre Verbreitung im digitalen Raum zu unterbinden. Dabei liegt der Schwerpunkt auf Desinformationen im nationalen Kontext, die durch private Akteure in der digitalen Öffentlichkeit lanciert werden.[4] Dies soll nicht darüber hinwegtäuschen, dass Desinformationskampagnen auch Teil von Propagandatätigkeiten anderer Staaten mit Wirkung auf die Bundesrepublik Deutschland sein können. Diese außen-

1 *Ingold*, in: Oppelland 2018, 85f.
2 S. *Holznagel*, MMR 2018, 18 (20).
3 S. *Häberle* 1995, 91.
4 S. zu Desinformationstätigkeit des Staates *Ingold* 2011.

und sicherheitspolitische Herausforderung erfordert insbesondere eine völkerrechtliche Analyse, auf die hier nur verwiesen werden kann.[5]

Zunächst wird aufgezeigt, wie Desinformationen rechtlich zu bewerten sind und welchen rechtlichen Schutz sie und besondere Verbreitungsformen wie Social Bots genießen und welche Regelungen uns andererseits vor Desinformationen schützen (5.A). Dies leitet zu der Frage über, welche Schutzpflichten und Handlungsspielräume des Staates im Hinblick auf Desinformationen im digitalen Raum bestehen (5.B). Anschließend wird die Aufdeckung und Bekämpfung von Desinformation durch staatliche Stellen und durch die Betreiber von Social Networks, die sich zu den wichtigsten Verbreitungskanälen von Desinformationen entwickelt haben,[6] untersucht (5.C). Den Abschluss bilden rechtspolitische Überlegungen zur Regulierung von Desinformationen (5.D).

A. Schutz von oder vor Desinformation nach geltendem Recht

Die Regelung von Desinformation ist eingebettet in verfassungsrechtliche Schutzaufgaben. Zum einen sind sowohl die Grundrechte betroffener Personen als auch die Bedingungen demokratischer Willensbildung gegenüber Desinformationen zu schützen. Zum anderen sind bei der Aufdeckung und Bekämpfung der Desinformation die in Art. 5 Grundgesetz (GG) verbürgten Kommunikationsgrundrechte zu beachten. Gefordert ist ein Ausgleich, der beide Schutzgewährleistungen miteinander in einen praktikablen Ausgleich bringt.

I. Schutz demokratischer Willensbildung

Sich eine Meinung zu bilden, sie mitzuteilen und über öffentliche Angelegenheiten zu diskutieren, ist Grundlage der Demokratie nach dem Grundgesetz. Art. 5 Abs. 1 GG beinhaltet daher in zweifacher Hinsicht grundlegende Gewährleistungen: Individualrechtlich wird mit der Meinungsfreiheit die Freiheit des Einzelnen zur Meinungsbildung, -äußerung und -verbreitung als Teil kommunikativer Selbstentfaltung erfasst. Zugleich schützt Art. 5 Abs. 1 GG den Prozess der Kommunikation und seine objektive Funktion für den

5 S. dazu *Thielbörger*, in: Oppelland 2018, 63.
6 *Lazer et al.*, Science 2018, 1094.

politischen Willensbildungsprozess.[7] Das Bundesverfassungsgericht formulierte schon im Lüth-Urteil aus dem Jahr 1958, dass die Meinungsfreiheit „für eine freiheitliche demokratische Staatsordnung schlechthin konstituierend" ist.[8]

Für die Teilhabe und Mitwirkung an öffentlichen Meinungsbildungsprozessen sind Informationen eine wichtige Voraussetzung.[9] Daher dient die ebenfalls von Art. 5 Abs. 1 Satz 1 GG verbürgte Informationsfreiheit dem Schutz der freien Meinungsbildung und schützt jeden, der sich aus allgemein zugänglichen Quellen informieren will.

Die besondere Funktion von Presse und Rundfunk für die individuelle und öffentliche Meinungsbildung wird in Art. 5 Abs. 1 GG verfassungsrechtlich anerkannt. Für die Presse entschied das Bundesverfassungsgericht schon vor über 50 Jahren, dass sie zur wahrheitsgemäßen Berichterstattung verpflichtet ist: Die Wahrheitspflicht wird nicht nur „um des Ehrenschutzes des Betroffenen willen gefordert", sondern „ist zugleich in der Bedeutung der öffentlichen Meinungsbildung im Gesamtorganismus einer freiheitlichen Demokratie begründet. Nur dann, wenn der Leser – im Rahmen des Möglichen – zutreffend unterrichtet wird, kann sich die öffentliche Meinung richtig bilden". Daher müssen Nachrichten und Behauptungen überprüft werden. Eine leichtfertige Weitergabe unwahrer Nachrichten ist unzulässig, erst recht eine bewusste Entstellung der Wahrheit.[10]

Die Meinungsbildung wird durch die inhaltliche Qualität der rezipierten Informationen geprägt. Auf diese nimmt das Recht im analogen und im digitalen Kommunikationsraum unterschiedlichen Einfluss. Die klassischen Massenmedien werden durch die Landespressegesetze erfasst und haben sich überwiegend dem Pressekodex unterworfen. Dagegen gibt es für Informationsangebote im Internet nur einen sehr schwachen rechtlichen Rahmen. Hier bestehen zahlreiche weitere (Des-)Informationsangebote, die von Blogs privater Einzelpersonen über Online-Angebote, die journalistische Methoden lediglich imitieren, bis hin zur elektronischen Presse reichen. In Social Networks sind all diese professionellen und nicht-professionellen Angebote verschiedenster Qualität und Ausrichtung verfügbar. Die Auswahl und Filterung der Inhalte und damit die Meinungsbildung werden sowohl durch die

7 BVerfGE 82, 272 (281) – Zwangsdemokrat; *Schmidt-Jortzig*, in: Isensee/Kirchhof 2009, § 162 Rn. 9.
8 BVerfGE 7, 198 (208) – Lüth.
9 BVerfGE 105, 252 (268) – Glykol; 105, 279 (301f.) – Osho; 113, 63 (77) – Junge Freiheit.
10 BVerfGE 12, 113 – Schmid-Spiegel.

Nutzenden selbst aber auch durch die Voreinstellungen und automatisierten Mechanismen der Netzwerkbetreiber wesentlich beeinflusst.

Im Gegensatz zu traditionellen Massenmedien können im Internet auch Einzelmeinungen eine beträchtliche Reichweite erfahren. Dem wurde bisher durch die Unterscheidung zwischen privater Meinungsäußerung und „journalistisch-redaktionellen" Angeboten Rechnung getragen. Dieser Begriff soll z. B. Blogs, die eine besondere meinungsbildende Funktion erfüllen, von anderen, etwa eher persönliches Erleben schildernden Angeboten, abgrenzen. Nur Telemedien mit journalistisch-redaktionell gestalteten Angeboten müssen gemäß § 54 Abs. 2 Rundfunkstaatsvertrag (RStV) den anerkannten journalistischen Grundsätzen entsprechen.[11]

Durch Desinformation kann der Prozess der individuellen und öffentlichen Meinungsbildung und damit die demokratische Willensbildung bedroht sein. Die öffentliche Kommunikation kann verzerrt sein, wenn Teilnehmer öffentlicher Diskurse eine nicht begründete Machtstellung erlangen (Paradigma kommunikativer Chancengleichheit).[12] Daher fordert der Schutz des Kommunikationsprozesses, die Vereinnahmung der öffentlichen Diskussion durch machtvolle Akteure zu unterbinden und Meinungsvielfalt zu gewährleisten.[13] Diese soll der sich selbst regulierende „Meinungsmarkt" erreichen. Entscheidend ist, dass er eine wesentliche Meinungsvielfalt und einen chancengerechten Zugang zur Teilhabe an Kommunikationsprozessen ermöglicht. Aus der Vielzahl an Meinungen, die als Basis der demokratischen Willensbildung dienen, soll sich die überzeugendste Meinung durchsetzen.[14]

Übertragen auf Desinformationen bedeutet dies, dass diese, abgesehen von festgesteckten Grenzen wie kollidierenden Persönlichkeitsrechten oder den Strafgesetzen, als Bedrohung des geschützten Kommunikationsprozesses zu sehen sind, wenn die Verbreiter von Desinformationen (wirtschaftliche) Machtstellungen innehaben, wenn sie in einer derart großen Anzahl auftreten, dass eine einseitige Informationslage gegeben ist oder wenn ein Informationsungleichgewicht vorliegt, das die übrigen gesellschaftlichen Kräfte nicht

11 S. BT-Drs. 17/12542, 50.

12 S. BVerfGE 20, 162 (176) – Augstein-Spiegel; 25, 256 (265) – Blinkfüer; *Hoffmann-Riem*, in: Benda u.a. 1995, § 7 Rn. 12, bezeichnet kommunikative „Privilegien" nach dem Grundsatz kommunikativer Chancengleichheit nicht als gänzlich ausgeschlossen, aber legitimierungsbedürftig; s. auch BT-Drs. 17/12542, 24.

13 BVerfGE 57, 295 (322); 73, 118 (LS 1 b); 57, 295 (320); 31, 314 (315); *Hartl* 2017, 35 m.w.N.; *Mengden* 2018, 71 ff.; s. grundsätzlich *Roßnagel,* in: EMR 2005, 35 ff.

14 S. BVerfGE 25, 256 (265) – Blinkfüer.

auszugleichen vermögen.[15] Von einer Bedrohung für die öffentliche Willensbildung kann daher insbesondere ausgegangen werden, wenn Desinformationen massenhaft verbreitet werden. Je mehr sie verbreitet werden, desto mehr sind sie zur Verzerrung der öffentlichen sowie politischen Meinungsbildung geeignet. Dann besteht auch umso mehr die Gefahr, Faktentreue oder zumindest das Streben nach dieser als gemeinsames Fundament der Kommunikation zu verlieren. Ein Verlust von Faktentreue kann außerdem die Gefahr begründen, dass Vertrauen in der Gesellschaft schwindet, das gemeinsames praktisches Wissen entstehen lässt und die Kooperationsfähigkeit in der Gesellschaft bewahrt.[16] Gerade für die politische Öffentlichkeit ist die Qualität der Informationen bedeutsam, da sie die Bürgerinnen und Bürger befähigen sollen, sachkundig an Wahlen und öffentlichen Diskussionen teilzunehmen.[17]

Spezifische Risiken für die demokratische Willensbildung gehen auch vom Einsatz von Malicious Bots, einer die freie Meinungsbildung gefährdenden Unterform von Social Bots,[18] aus: Da sie in der Regel nicht als Computerprogramme erkennbar sind, können die Nutzenden von Malicious Bots in einer Weise in die öffentliche Meinungsbildung eingreifen, wie dies bisher nicht möglich war.[19] Ihre Programmierung lässt es zu, dass sie z. B. zigtausende Nutzerprofile „bespielen" können, indem sie vorprogrammierte Beiträge posten, auf Basis einer bestimmten Verschlagwortung andere Beiträge wiederholen oder eigene Beiträge verfassen.[20] Auch können sie den Eindruck erwecken, hinter einer Meinung stünden sehr viele verschiedene Individuen. Damit sind sie ein geeignetes Instrument, um Falschnachrichten schnell zu verbreiten und die öffentliche Meinungsbildung zu verzerren.[21]

Die von Malicious Bots manipulierten Stimmungsbilder können ihre Reichweite erheblich erhöhen, wenn sie als solche unerkannt in „etablierte" Medienöffentlichkeiten wie Tageszeitungen und ihre Online-Ableger sowie TV-Berichterstattung eingehen oder Grundlage von Big-Data-Analysen werden.[22] Insbesondere in Bezug auf singuläre Ereignisse wie Wahlen oder Krisensituationen könnte eine aktive Stimmungsmache durch Malicious Bots

15 BVerfGE 105, 252 (268) – Glykol.
16 *Ladeur*, in: Eifert/Gostomzyk 2018, 169 (182f.).
17 *Holznagel*, NordÖR 2011, 205 (209).
18 S. zu diesen Kap. 1.C und 4.B.5.
19 S. hierzu auch *Löber/Roßnagel*, MMR 2019, 493 (494).
20 S. dazu *Freitas et al.*, Social Network Analysis and Mining 2016, 1 ff.
21 S. *Ross et al.*, EJIS 2019, 1; s. auch BT-Drs. 19/6970, 15 ff.
22 *Dankert/Dreyer*, K&R 2017, 73 (78); *Dankert*, in: Hoffmann-Riem 2018, 160 ff.; *Kind u.a.* 2017, 66.

zu erheblichen Verwerfungen führen. Hier könnten sie kurzzeitig sehr wirkmächtig werden, wenn die Manipulation erst aufgedeckt wird, wenn das Ereignis schon wieder vorbei ist.[23]

Malicious Bots sind zwar nur ein Instrument unter mehreren, um die Verbreitung von Desinformationen zu beschleunigen – diese werden derzeit noch vor allem durch interessierte Menschen verbreitet.[24] Sie sind jedoch ein ideales Instrument, um das öffentliche Meinungsklima zu beeinflussen. Nach der Theorie der „Schweigespirale" orientieren sich Menschen am Handeln anderer, fürchten sich vor sozialer Isolation und davor, sich gegen ein als mehrheitlich wahrgenommenes Meinungsbild zu stellen und tendieren daher dazu, sich der Mehrheitsmeinung sogar anzuschließen.[25] In einer Simulationsstudie konnte gezeigt werden, dass bereits eine relativ geringe Anzahl von Malicious Bots (zwei bis vier Prozent der Meinungsbekundungen) in einem Netzwerk das Meinungsklima in zwei Dritteln der Fälle zu deren Gunsten verändern kann.[26] Durch massenhaften Einsatz von automatisierter zielgerichteter Kommunikation, wie sie mittels Malicious Bots möglich ist, wird die freie demokratische Willensbildung gefährdet.[27]

Darüber hinaus birgt die Verwendung von Malicious Bots erhebliche Risiken für die Integrität menschlicher Kommunikation. Malicious Bots können zu sinkendem Vertrauen in menschliche Kommunikation beitragen. Sie treten in Social Networks auf wie andere Menschen. Sie sind so programmiert, dass sie vortäuschen, ein menschlicher Kommunikationspartner zu sein. Die Kommunikation zwischen Menschen ist die grundlegende Voraussetzung für die freie Entwicklung und Entfaltung der jeweiligen Persönlichkeit, die Entwicklung einer eigenen Meinung und auch für die gleichberechtigte Teilnahme an der öffentlichen Willensbildung sowie die Vorbereitung demokratischer Entscheidungen. Diese grundlegenden Voraussetzungen von Autonomie und Demokratie sind nur gewährleistet, wenn die menschliche Kommunikation nicht technisch manipuliert ist und wenn die Teilnehmer in die Integrität der Kommunikation ihres Gegenübers vertrauen können. Werden sie aber im Netz immer häufiger mit Bots konfrontiert, die sich mit ihnen in natürlicher

23 *Hegelich* 2016, 7; ein erkennbares Risiko verneinen *Schulz/Dreyer* 2018, 17.

24 *Vosoughi/Roy/Aral*, Science 2018, 1146; s. *Schwenkenbecher*, „So verbreiten sich falsche Nachrichten im Internet", Süddeutsche Zeitung vom 8.3.2018.

25 Näher zur Schweigespirale *Noelle-Neumann* 1996, 59 ff.; s. auch *Graber/Lindemann*, in: Sachs-Hombach/Zywietz 2018, 59 ff.

26 *Ross et al.,* EJIS, 2019, 1; s. auch die Untersuchung von 4,4 Mio. Twitter-Tweets von *Kušen/Strembeck*, Online Social Networks and Media, Vol. 10-11 (2019), 1 ff.

27 S. *Löber/Roßnagel*, MMR 2019, 493 (494).

Sprache so unterhalten, dass sie nicht mehr unterscheiden können, ob dies ein Algorithmus oder ein Mensch ist, können menschliche Kommunikationsprozesse und mit ihnen Autonomie und Demokratie gewichtige Verluste erleiden. Diese werden erheblich zunehmen, wenn in naher Zukunft zusätzlich die Möglichkeiten künstlicher Intelligenz und selbstlernender Systeme zur Steigerung des Täuschungseffekts genutzt werden.

Bei Social Bots überschneiden sich die Risiken für die öffentliche Meinungsbildung und die Risiken für eine integre und vertrauenswürdige Kommunikation zwischen Menschen. Das mit Social Bots verbundene Problem würde nicht richtig wahrgenommen, wenn die (rechts-)politische Diskussion es auf den einen oder den anderen Aspekt beschränkt. Vielmehr muss sie beide Risikodimensionen gemeinsam sehen – und für beide gemeinsam mögliche (rechts-)politische Vorsorgemöglichkeiten erörtern.

II. Schutz der Meinungs- und Informationsfreiheit

Für die verfassungsrechtliche Bewertung von Desinformationen ist aber auch der grundrechtlich gewährleistete Schutz der Meinungs- und Informationsfreiheit zu beachten. Für die Frage, ob eine Äußerung unter den Schutz der Meinungsfreiheit fällt, ist zwischen Werturteil und Tatsachenbehauptung zu unterscheiden:[28] Meinungen sind geprägt „durch das Element der Stellungnahme, des Dafürhaltens oder Meinens“ und als subjektive Ansichten nicht beweisbar.[29]Tatsachenbehauptungen werden durch die objektive Beziehung zwischen der Äußerung und der Realität charakterisiert. Sie sind, anders als Meinungen, einer Überprüfung auf ihren Wahrheitsgehalt, also dem Beweis, zugänglich.[30] Bloße Tatsachenbehauptungen fallen nicht unter die Meinungsfreiheit.[31] Soweit die Tatsachenbehauptungen Dritten zur Meinungsbildung dienen sollen, fallen sie dennoch in den Schutzbereich der Meinungsfreiheit.[32] Ist die Trennung der tatsächlichen und wertenden Bestandteile einer Äußerung nicht möglich, ist von einer Meinungsäußerung auszugehen.[33]

Der Begriff der Meinung ist weit zu verstehen. Auch scharfe, polemische, provokative oder abstoßende Meinungsäußerungen fallen in den Schutzbe-

28 Ausführlich hierzu *Rühl*, AfP 2000, 17.
29 BVerfGE 85, 1 (14); 61, 1 (9) – NPD Europas.
30 St. Rspr. BVerfGE 65, 1 (41) – Volkszählung; 90, 241 (247) – Auschwitzlüge.
31 BVerfGE 61, 1 (8f.) – NPD Europas.
32 BVerfGE 85, 1 (15); 54, 208 (219); 61, 1 (8) – NPD Europas.
33 BVerfG, BeckRS 2013, 54173, Rn. 18.

reich.[34] Auf den Wert oder Unwert einer Meinung kommt es ebenso wenig an wie auf die Qualität der einer Meinung zugrundeliegenden Quellen, da die Kommunikation nicht um ihres Inhalts, sondern um ihrer selbst willen geschützt wird.[35] „Denn das Grundrecht der Meinungsfreiheit will nicht nur der Ermittlung der Wahrheit dienen; es will auch gewährleisten, dass jeder frei sagen kann, was er denkt, auch wenn er keine nachprüfbaren Gründe für sein Urteil angibt oder angeben kann."[36]

Zugleich soll die Meinungsfreiheit auch der Wahrheitsfindung im Kommunikationsprozess dienen. So sind bewusst unwahre Tatsachen (bewusste Lüge) und Tatsachen, deren Unwahrheit im Zeitpunkt der Äußerung unzweifelhaft feststeht, vom Schutzbereich ausgeschlossen.[37] Sie können keinen sinnvollen Beitrag zur verfassungsrechtlich vorausgesetzten Aufgabe zutreffender Meinungsbildung leisten.[38] Bei den Mischäußerungen, in denen Meinung und Tatsachenbehauptung untrennbar miteinander verbunden sind, ist jedoch im Interesse eines effektiven Grundrechtsschutzes die Äußerung, selbst wenn der tatsachenbasierte Äußerungsteil erwiesen unwahr ist, insgesamt vom Schutzbereich der Meinungsfreiheit erfasst.[39] Dies kann auch auf Desinformationen zutreffen, wenn die bewusst unwahren Tatsachenbehauptungen z. B. in Form von vermeintlich wissenschaftlichen Belegen oder erfundenen Zitaten untrennbar mit Werturteilen verbunden sind.

An die Sorgfaltspflicht dürfen keine zu hohen Anforderungen gestellt werden, „die die Bereitschaft zum Gebrauch des Grundrechts herabsetzen und so auf die Meinungsfreiheit insgesamt einschnürend wirken können".[40] Dies wird vor allem relevant für Tatsachenbehauptungen, deren Unwahrheit im Zeitpunkt der Äußerung nicht zweifelsfrei feststeht. Wahrheit oder Unwahrheit werden oft erst im Prozess der Kommunikation ersichtlich.[41] Für die Aufnahme in den Schutzbereich ist nicht nötig, dass Recherchen zur Überprüfung des Wahrheitsgehalts durchgeführt wurden.

---

34 BVerfG, NJW 2014, 3357 (3358), Rn. 11; BVerfGE 61, 1 (9f.) – NPD Europas.
35 *Fechner*, in: Stern/Becker 2018, Art. 5 GG Rn. 80f.
36 BVerfGE 42, 163 (171) – Deutschland-Stiftung.
37 BVerfGE 61, 1 (8) – NPD Europas; 54, 208 (219); 99, 185 (197).
38 BVerfGE 54, 208 (219); 61, 1 (8) – NPD Europas; 99, 185 (197); 114, 339 (352f.) – Stolpe; s. *Bethge*, in: Sachs 2018, Art. 5 GG Rn. 28; *Grimm*, NJW 1995, 1697 (1699); *Steinbach*, JZ 2017, 653 (656).
39 BVerfG, NJW-RR 2017, 1001 m.w.N.; BVerfG, NJW 2012, 1498 (1499); *Ulrich*, AfP 2017, 316.
40 BVerfGE 114, 339 (353) – Stolpe m.w.N.
41 *Steinbach*, JZ 2017, 653 (655).

Anonymität erschwert die Aufdeckung der Verursacher von Desinformation. Sie begünstigt die Verbreitung von Desinformationen, „Shitstorms", Hassreden sowie Social Bots.[42] Allerdings ist eine anonyme Meinungsäußerung grundrechtlich geschützt.[43] Sie ist ein Ausdruck der Grundrechte auf informationelle Selbstbestimmung und auf Meinungsfreiheit. Zu letzterem gehört das Recht, die Umstände der Meinungsäußerung frei zu wählen. Eine Beschränkung des Schutzbereichs auf identifizierte Äußerungen würde die Gefahr der Selbstzensur begründen und ist deshalb mit dem Grundrecht nicht vereinbar.[44] Anonymität und auch Pseudonymität sollen den sich Äußernden vor etwaiger Voreingenommenheit gegen seine Person schützen und es ihm ermöglichen, frei zu sprechen, ohne Sanktionen oder sonstige negative Auswirkungen befürchten zu müssen.[45] Diese Erleichterung kann auch förderlich für den freien Kommunikationsprozess sein, indem etwa Anhänger von Mindermeinungen eher von ihrer Äußerungsfreiheit Gebrauch machen und somit zu einem vielfältigen Diskurs beitragen.

Auch *Social Bots* können vom Schutzbereich der Meinungsfreiheit umfasst sein.[46] Sie enthält auch das Recht, die Modalitäten der Äußerung und das Verbreitungsmedium frei zu wählen.[47] Dies gilt auch für neue technische Artikulations- und Verbreitungsmittel.[48] Dazu zählt grundsätzlich auch die Freiheit, durch Bots unter Pseudonymen eine Meinung zu verbreiten.

Keinen Schutz verdient jedoch die Äußerungsform, dass der Bot die Identität einer anderen, real existierenden Person übernimmt.[49] Sie enthält die bewusst unwahre Tatsachenbehauptung, eine andere natürliche Person habe diese Aussage getätigt (bewusstes Falschzitat). Wenn so einem anderen Grundrechtsträger Äußerungen in den Mund gelegt werden, die er tatsächlich nicht getroffen hat, verletzt dies dessen allgemeines Persönlichkeitsrecht „in besonderem Maße".[50]

---

42 *Lausen*, ZUM 2017, 278 (288); *Paal/Hennemann*, JZ 2017, 641 (644); *Schliesky u.a.* 2014, 125f.
43 S. *Kersten*, JuS 2017, 193 (195).
44 So die h.M. BGHZ 181, 328 – spickmich; näher dazu *Heilmann* 2013, 98 ff. m.w.N.; s. auch *Ballhausen/Roggenkamp*, K&R 2008, 403 (406); *Starck/Paulus*, in: v. Mangoldt/Klein/Starck 2018, Art. 5 GG Rn. 92.
45 BGHZ 181, 328 – spickmich.
46 S. *Löber/Roßnagel*, MMR 2019, 493 (496).
47 *Jarass*, in: Jarass/Pieroth 2018, Art. 5 GG Rn. 9; *Steinbach*, ZRP 2017, 101 (102).
48 *Bethge*, in: Sachs 2018, Art. 5 GG Rn. 44.
49 S. *Löber/Roßnagel*, MMR 2019, 493 (496).
50 BVerfGE 54, 208 (219f.).

Keinen Schutz verdient auch der Einsatz von Bots, um massenhaft Beiträge zu generieren und damit ein verfälschtes Meinungsbild und „trending topics" zu erzeugen.[51] Diese „multiple Identitätskreierung"[52] stellt eine bewusste Täuschung über die Anzahl der Äußernden dar. Sie suggeriert, dass sehr viele menschliche Nutzerinnen und Nutzer hinter der geäußerten Meinung stehen und ein gesteigertes (gesamt-)gesellschaftliches Interesse an dieser Position besteht.[53] Diese Beeinflussung des Meinungsbildungsprozesses vermag die Schweigespirale auszulösen und Mehrheitsverhältnisse zu verändern. Aufgrund der funktionalen Ausrichtung der Meinungsfreiheit auf den demokratischen Willensbildungsprozess kann die Täuschung darüber, dass eine Meinung nicht von einer Person, sondern von vielen Tausendenden Mitbürgerinnen und Mitbürgern vertreten wird, nicht von der Meinungsfreiheit geschützt sein. Dies verstößt gegen den demokratischen Grundsatz „one man, one voice".[54] Zwar können in Social Networks auch echte Nutzerinnen und Nutzer ihre Meinung manuell ohne größeren Aufwand immer wieder erneut kundtun, etwa mithilfe von Copy-and-Paste. Dadurch wird aber die Meinung nur eines Menschen mehrfach wiederholt. Durch den Einsatz von Bots täuscht der Verwendende jedoch vor, hinter jeder Äußerung stehe ein realer Mensch und verzerrt damit das Meinungsbild zugunsten seiner Meinung. Soweit die Bot-Nutzung nicht unter die Meinungsfreiheit fallen sollte, gilt für sie jedoch die allgemeine Handlungsfreiheit nach Art. 2 Abs. 1 GG.[55]

Die Kommunikationsfreiheiten sind nicht vorbehaltlos gewährleistet, sondern unterliegen nach Art. 5 Abs. 2 GG den *Schranken*, die sich aus den Vorschriften der allgemeinen Gesetze sowie den gesetzlichen Bestimmungen zum Schutze der Jugend und dem Recht der persönlichen Ehre ergeben.[56] Allgemein ist ein Gesetz, wenn es nicht gegen einen spezifischen Meinungsinhalt gerichtet ist. Zum Schutz der Jugend oder der Ehre sind Beschränkungen der Meinungsfreiheit nur zulässig, wenn sie bei einer Abwägung zwischen den widerstreitenden Rechtspositionen gewichtiger sind als der Schutz der Meinungsfreiheit. Eine solche Abwägung erübrigt sich allerdings, wenn es sich bei der Äußerung um Schmähkritik handelt.[57]

---

51 S. *Kind u.a.* 2017, 56f.; *Freitas et al.*, Social Network Analysis and Mining 2016, 1 (15) m.w.N.
52 *Steinbach*, ZRP 2017, 101 (103).
53 *Brings-Wiesen* 2016.
54 S. auch *Schröder*, DVBl 2018, 465 (468).
55 S. *Löber/Roßnagel*, MMR 2019, 493 (496).
56 S. näher zur Schrankentrias statt vieler *Bethge*, in: Sachs 2018, Art. 5 GG Rn. 136 ff.
57 BVerfG, NJW 2017, 1460 (1460f.).

Meinungsäußerungen, die im Wesentlichen unwahre Tatsachenbehauptungen enthalten und in das *allgemeine Persönlichkeitsrecht* Dritter eingreifen, sind in der Regel rechtswidrig. Denn bei Äußerungen, in denen sich wertende und tatsächliche Elemente in der Weise vermengen, dass die Äußerung insgesamt als Meinungsäußerung anzusehen ist, fällt bei der Abwägung maßgeblich der Wahrheitsgehalt der tatsächlichen Bestandteile ins Gewicht.[58] Wahre Tatsachenbehauptungen sind in der Regel hinzunehmen, auch wenn sie nachteilig für die Betroffenen sind.[59] Auch kommt einer Äußerung umso größerer Schutz zu, je mehr sie sich als Werturteil denn als Tatsachenbehauptung einordnen lässt.[60] Wenn die Meinungsäußerung jedoch einen erwiesen falschen oder bewusst unwahren Tatsachenkern enthält oder die mit ihr verbundene und ihr zugrundeliegende Tatsachenbehauptung erwiesen unwahr ist, tritt das Grundrecht der Meinungsfreiheit regelmäßig hinter die Schutzinteressen des von der Äußerung Betroffenen zurück.[61]

In den Fällen, in denen die Unwahrheit im Zeitpunkt der Äußerung noch nicht feststeht, wendet die Rechtsprechung differenzierte Sorgfaltsstandards an.[62] Bei Einhaltung der Sorgfaltspflichten genießt auch die sich im Nachhinein als falsch erweisende Tatsachenbehauptung Grundrechtsschutz.[63] „Journalistischen Laien" obliegen dabei geringere Sorgfaltspflichten als der Presse (sog. Laienprivileg).[64]

Desinformation ist von *Satire* zu unterscheiden, die grundsätzlich von der Meinungs- und Kunstfreiheit[65] gedeckt ist. Auch satirische Inhalte enthalten nicht selten bewusst unwahre Tatsachenbehauptungen. Im Unterschied zu Desinformationen werden sie jedoch nicht von einer Täuschungs- und Manipulationsabsicht getragen und sind in der Regel als solche erkennbar. Dies gilt nicht für Desinformationen, die versuchen mit einem kleinen, schwer

58 BGH, MDR 2016, 648 Rn. 51; BVerfG, NJW 2012, 1643 Rn. 34; BVerfG, NJW 2013, 217 (218); BVerfG, AfP 2009, 480 Rn. 62.

59 Begrenzungen ergeben sich, wenn z. B. wahre Tatsachen aus der Intimsphäre und engsten Privatsphäre einer Person verbreitet werden, s. *Starck/Paulus*, in: v. Mangoldt/Klein/Starck 2018, Art. 5 GG Rn. 329 ff.

60 S. BVerfGE 61, 1 – NPD Europas; *Woger/Männig*, PinG 2017, 233 (235).

61 BGH, MDR 2016, 648 Rn. 51; BVerfGE 90, 241 (248f.) – Auschwitzlüge; BVerfG, NJW 2012, 1643 Rn. 33f.

62 *Steinbach*, JZ 2017, 653 (656).

63 BGH, AfP 2017, 316 Rn. 24; *Rühl*, AfP 2000, 17.

64 BVerfGE 99, 185 (197f.); BVerfG, AfP 2009, 480 Rn. 62 m.w.N.; BVerfGE 85, 1 (15 ff.).

65 BVerfGE 30, 173 – Mephisto; 75, 369 – Strauß-Karikaturen; s. auch *Faßbender*, NJW 2019, 705.

auffindbaren Hinweis auf „Satire“ etwa verleumderische Behauptungen unter dem Deckmantel der Satire zu verbreiten.

Im Bewusstsein ihrer Unwahrheit verfasste Aussagen können auch nicht durch die *Wissenschaftsfreiheit* gemäß Art. 5 Abs. 3 Satz 1 GG gedeckt sein. Denn unter Wissenschaft ist jede Tätigkeit zu fassen, die „nach Inhalt und Form als ernsthafter und planmäßiger Versuch zur Ermittlung der Wahrheit anzusehen ist“.[66] Auch liegt keine Wissenschaft vor, wenn das Verhalten von der Absicht getragen ist, ein Handeln anderer auszulösen, etwa bei politischen, weltanschaulichen Aktionen oder „ideologischer Indoktrination“.[67] Das ernsthafte Bemühen um Erkenntnis muss möglich sein – auch wenn sich später herausstellen sollte, dass der Äußernde mit seiner Annahme falsch lag. Der Beweis des Gegenteils durch neuere Erkenntnisse ist gerade ein Teil der wissenschaftlichen Forschung.[68] Wissenschaftliche Erkenntnisse können dazu beitragen, Desinformationen den Boden zu entziehen, wenn sie in nachprüfbarer Weise die bewusst unwahre Tatsachenbehauptungen widerlegen.

Im einfachen Gesetzesrecht steht gegen Desinformation ein vielfältiges Instrumentarium zur Verfügung. Es reicht von zivilrechtlichen Unterlassungsansprüchen über Straftatbestände hin zu speziellen lauterkeitsrechtlichen Irreführungsverboten und speziellen Anforderungen für journalistisch-redaktionelle Medien.

III. Schutz im Zivilrecht

Zivilrechtlich schützt § 823 Abs. 1 Bürgerliches Gesetzbuch (BGB) das allgemeine Persönlichkeitsrecht. Weisen unwahre Tatsachenbehauptungen einen konkreten Personenbezug auf und greifen ungerechtfertigt in das allgemeine Persönlichkeitsrecht ein, kann der Betroffene nach §§ 1004, 823 BGB analog Beseitigung und Unterlassung der rechtsverletzenden Äußerungen und ihrer Verbreitung erreichen. Diesen Anspruch kann er auch im Wege des einstweiligen Rechtsschutzes verfolgen. Ansprüche auf Gegendarstellung der unwahren Tatsachenbehauptung bestehen hingegen bei digitalen Desinformationen häufig nicht, da diese nach § 56 RStV nur bei der Verbreitung über journalistisch-redaktionell gestaltete Telemedien geltend gemacht

66 BVerfGE 35, 79 (113); 47, 327 (367); 90, 1 (12).
67 *Scholz*, in: Maunz/Dürig 2019, Art. 5 GG Rn. 93 m.w.N.
68 *Fritzsche*, in: Spickhoff 2018, § 3 HWG Rn. 6.

werden können.[69] Faktisch sind diese Ansprüche jedoch in Anbetracht anonymer Äußerungen sowie leichter Reproduzier- und Verlinkbarkeit von Veröffentlichungen nur schwer durchsetzbar.

IV. Schutz im Medienrecht

Werden Falschinformationen über das Internet, rechtlich Telemedien, verbreitet, gelten die strengsten Voraussetzungen, wenn es sich um journalistisch-redaktionelle Angebote handelt. Dann sind gemäß § 54 Abs. 2 Satz 2 RStV besondere Sorgfalts- und Recherchepflichten zu beachten. Anbieter von Social Networks fallen nach herrschender Meinung nicht hierunter. Sie machen sich in der Regel die Inhalte nicht zu eigen und tragen keine journalistisch-redaktionelle Verantwortung.[70] Diese Pflichten der elektronischen Presse gelten jedoch für Webseiten, die darauf ausgerichtet sind, am Prozess der öffentlichen Meinungsbildung teilzunehmen, deren Angebote journalistisch gestaltet sind und die eine gewisse Kontinuität und Dauerhaftigkeit aufweisen.[71] Sie gelten auch dann, wenn diese Angebote journalistische Methoden lediglich imitieren. Daher können neben professionellen Angeboten auch Laienangebote zur elektronischen Presse gehören.[72] Obwohl die bewusste Verbreitung von Falschinformationen einen besonders schweren Verstoß gegen die anerkannten journalistischen Grundsätze und die Sorgfaltspflichten darstellt, schließt § 59 Abs. 3 RStV Aufsichtsmaßnahmen der Landesmedienanstalten bei Verstößen gegen § 54 RStV aus. Der Deutsche Presserat ist zwar auch für journalistisch-redaktionelle Medien zuständig, kann eine öffentliche Rüge im Rahmen der Selbstregulierung aber nur dann verhängen, wenn sich die veröffentlichende Stelle diesem durch Selbstverpflichtung unterworfen hat.[73] Eine selbstverpflichtende Bekennung zum Pressekodex und damit die Begründung der Zuständigkeit des Presserats liegt jedoch bei

69 *Koreng*, KriPoZ 2017, 151 (156); *Petruzzelli*, MMR 2017, 800 (802).

70 *Paal*, MMR 2018, 567 (569); *Peifer*, CR 2017, 809 (812); a.A. *Mengden* 2018, 311f. für eine Anwendbarkeit von § 54 RStV auf algorithmengestützte Zugangsdienste.

71 Die klassischen Bestandteile Universalität, Aktualität, Periodizität und Publizität der Publizistikdefinition können in abgewandelter Form auch für die Auslegung des Online-Journalismus fruchtbar gemacht werden, *Lent*, in: Gersdorf/Paal 2019, § 54 RStV Rn. 5 m.w.N.

72 *Lent*, in: Gersdorf/Paal 2019, § 54 RStV Rn. 5.

73 S. *Holznagel*, MMR 2018, 18 (21).

Online-Angeboten, die vermehrt Desinformationen verbreiten, zumeist nicht vor.

V. Schutz im Strafrecht

Desinformation kann unter das Strafrecht fallen.[74] Die Veröffentlichung von allgemeinen Falschnachrichten ohne Bezug zu einer bestimmten Person oder Personengruppe ist in der Regel nicht strafbar. Ehrschutzdelikte kommen in Betracht, wenn Desinformationen ehrenrührige Tatsachenbehauptungen enthalten. Ehrverletzenden Äußerungen im digitalen Raum kommt oftmals ein erhöhter Unrechtsgehalt zu, der aus den Besonderheiten des „Tatmittels Internet" resultiert, namentlich der Ubiquität, permanenten Verfügbarkeit und Nicht-Eliminierbarkeit sowie ihrer schnellen Verbreitung.[75] Für eine Verleumdung ist gemäß § 187 Strafgesetzbuch (StGB) erforderlich, dass der Täter wider besseres Wissen, das heißt in sicherer Kenntnis der Unwahrheit handelte,[76] während die üble Nachrede gemäß § 186 StGB bei fehlendem Wissen hinsichtlich der Unwahrheit der verbreiteten Tatsache, die zudem nicht erweislich wahr sein muss, in Betracht kommt. Somit kann die üble Nachrede auch bei „gutgläubigen" Tätern erfüllt sein, die Beiträge in Social Networks teilen und denen dabei nicht bekannt ist, dass es sich um eine Desinformation handelt.[77]

Die strafbaren Äußerungen können sich nicht nur auf einzelne Personen, sondern auch auf Kollektive oder Individuen als Mitglieder der Kollektive beziehen. Der Anwendungsbereich der Ehrschutzdelikte ist jedoch in Fällen von pauschalen Äußerungen gegen bestimmte Bevölkerungsgruppen erheblich eingeschränkt. Die sog. Kollektivbeleidigung, die auch in Form von Verleumdungen und übler Nachrede auftreten kann, erfordert das Vorliegen folgender Bedingungen: Erstens muss die Personengesamtheit eine rechtlich anerkannte gesellschaftliche (auch wirtschaftliche) Funktion erfüllen und zweitens muss sie einen einheitlichen Willen bilden können.[78] Die

74 S. zusätzlich auch §§ 100a, 109d, 241a, 145d, 153 ff., 164f., 263, 269, 353a StGB.
75 *Hilgendorf*, ZIS 2010, 208 (213); *ders*., EWE 2008, 403, (409); *Krischker*, JA 2013, 488 (489).
76 *Kühl*, in: Lackner/Kühl 2018, § 187 StGB Rn. 1.
77 Näher zur Strafbarkeit der Weiterverbreitung in Social Networks *Krischker*, JA 2013, 488 ff.
78 BGHSt 6, 186 (191f.); näher zu Hate Speech im Internet *Koreng*, KriPoZ 2017, 151 (153 ff.).

davon zu unterscheidende Beleidigung unter einer Kollektivbezeichnung setzt hingegen eine ehrverletzende Äußerung in Bezug auf eine hinreichend überschaubare und abgegrenzte Personengruppe voraus.[79] Wenn sich die falsche Meldung etwa auf „Flüchtlinge", „Homosexuelle", „die Politiker" oder Menschen einer bestimmten Nationalität bezieht, werden diese Voraussetzungen in der Regel nicht erfüllt sein, da keine Betroffenheit einzelner Gruppenmitglieder bejaht werden kann.[80]

Außerdem können digitale Desinformationen insbesondere Straftatbestände erfüllen, die (auch) dem Schutz des öffentlichen Friedens dienen. Nach § 130 StGB ist eine Volksverhetzung strafbar, wenn durch grob wahrheitswidrige und einseitige Verzerrungen Hass gegen von der Vorschrift geschützte Bevölkerungsgruppen geschürt wird. Wenn die entsprechende Falschmeldung sowohl in qualitativer als auch in quantitativer Hinsicht zur Friedensstörung geeignet ist, kann § 130 Abs. 1 StGB einschlägig sein. Doch auch wenn diese recht hohe Schwelle nicht überschritten wird, kann die Desinformation gemäß Abs. 2 strafbar sein, wenn sie über das Internet und mithin über ein „Telemedium" verbreitet wird.[81]

Desinformation kann unter § 126 Abs. 2 StGB (Störung des öffentlichen Friedens) fallen, wenn wider besseres Wissen vorgetäuscht wird, eine Katalogtat aus Abs. 1 (Mord, Totschlag u.a.) stünde bevor. Bejaht wurde dies bereits für einen Blogeintrag über einen erfundenen Terroranschlag. Erforderlich ist hierfür insbesondere, dass die Veröffentlichung des Artikels geeignet ist, „Teile der Bevölkerung bzw. eine nicht unbeträchtliche Personenmehrheit ernsthaft zu beunruhigen und das Vertrauen in die öffentliche Rechtssicherheit zu beeinträchtigen".[82]

Im Hinblick auf den Schutz politischer Wahlen sehen §§ 107 ff. StGB Straftatbestände vor. Diese werden durch Desinformation im Vorfeld von Wahlen indes regelmäßig nicht erfüllt sein. Eine Wählertäuschung nach § 108a StGB liegt noch nicht vor, wenn jemand durch falsche Wahlpropaganda veranlasst wird, in einem bestimmten Sinn oder gar nicht zu wählen.[83]

79 S. BGH, NJW 2017, 1092 Rn. 16f.; BGH, NJW 2016, 2643 Rn. 16f.
80 *Koreng*, KriPoZ 2017, 151 (153f.).
81 Eingehend dazu *Hoven/Krause*, JuS 2017, 1167 (1169f.).
82 AG Mannheim, MMR 2019, 341 Rn. 35.
83 *Eser*, in: Schönke/Schröder 2019, § 108a StGB Rn. 2.

B. Handlungspflichten und -möglichkeiten des Staats

Im Spannungsfeld zwischen dem Schutz des öffentlichen Diskurses und der demokratischen Willensbildung einerseits und dem Schutz der Kommunikationsgrundrechte andererseits besteht für die staatlichen Organe bisher ein Handlungsrahmen, der für die neuen Herausforderungen von Desinformationen und Manipulationsmöglichkeiten wie Malicious Bots genutzt werden kann, der aber nicht für sie entwickelt worden ist und sie auch nicht vollständig abdeckt. Insbesondere den Gesetzgeber trifft jedoch die generelle Verpflichtung, eine funktionierende Kommunikationsordnung zu gewährleisten, um den Bürgern die Wahrnehmung ihrer Freiheiten zu ermöglichen.[84] Daher ist der Frage nachzugehen, inwieweit ein objektiv-rechtlicher Auftrag aus den Schutzaufgaben an den Gesetzgeber abzuleiten ist, über den bestehenden Ordnungsrahmen hinaus den öffentlichen und demokratischen Diskurs in digitalen Räumen vor Desinformationen und verzerrten Stimmungsbildern zu schützen.

Grundsätzlich soll der Meinungsbildungsprozess ohne staatliche Einmischung erfolgen.[85] Die politische Willensbildung soll sich vom Volk zu den Staatsorganen hin und nicht umgekehrt vollziehen.[86] Der Umgang mit Desinformationen ist somit grundsätzlich der gesellschaftlichen Sphäre zugewiesen.[87] Der Verbreitung von Unwahrheiten soll in erster Linie der Meinungskampf entgegenwirken, indem in ihm Desinformationen aufgedeckt und richtiggestellt werden.[88] Nur in besonderen Bedrohungslagen von einer gewissen Qualität, in denen die bereitstehenden Mittel des Meinungskampfs nicht genügen, um unzulässige Beeinflussungen abzuwehren, sind staatliche Maßnahmen in Erwägung zu ziehen. Von einer Handlungspflicht des Gesetzgebers ist daher erst bei einer unmittelbaren Beeinträchtigung und ernsthaften Bedrohung der öffentlichen Meinungsbildung auszugehen.[89] Diese könnte etwa bei einer „existenziellen Gefährdung des normativen Vielfaltsleitbildes" anzunehmen sein.[90] Ob eine derartige Gefährdung besteht, hängt somit von

84 *Kühling*, in: Gersdorf/Paal 2019, Art. 5 GG Rn. 13 m.w.N.; *Schulze-Fielitz*, in: Dreier 2013, Art. 5 Abs. 1, 2 GG Rn. 218 ff.; *Schemmer*, in: Epping/Hillgruber 2019, Art. 5 GG Rn. 1 m.w.N.; *Bull*, Der Staat (58) 2019, 57 (93).
85 BVerfGE 20, 56 (99) – Parteienfinanzierung I; 78, 350 (363).
86 S. BVerfGE 20, 56 (99) – Parteienfinanzierung I; s. auch *Hillgruber*, JZ 2016, 495 (498).
87 Ebenso *Ingold*, in: Oppelland 2018, 101.
88 *Milker*, ZUM 2017, 216 (220).
89 *Libertus,* ZUM 2018, 20 (22) m.w.N.
90 *Dankert/Dreyer*, K&R 2017, 73 (75).

der Einschätzung ihrer Wahrscheinlichkeit und ihres Schadenspotenzials für den demokratischen Prozess der freien und gleichberechtigten politischen Willensbildung ab. Hält man diesen für unmittelbar gefährdet, muss der Gesetzgeber diesen durch geeignete Maßnahmen schützen.

Doch auch wenn diese Gefährdung verneint wird, ist der Gesetzgeber im Rahmen seiner Einschätzungsprärogative befugt, Regelungen zu erlassen, die diesem Risiko vorbeugen und zur Aufrechterhaltung einer funktionierenden Kommunikationsordnung beitragen.[91] Gerade im Hinblick auf Malicious Bots gibt es ausreichende Anhaltspunkte für die Erwartung, dass ihnen in Zukunft ein größeres Gewicht zukommen wird.[92] Bereits heute liegen die technischen Voraussetzungen für einen „großflächigen Einsatz von Social Bots in Form von Bot-Armeen" vor.[93] Auch die Relevanz von Social Networks als Informationsquelle für die Meinungsbildung könnte weiter zunehmen.[94]

Ein Gesetz, das zum Schutz integrer Kommunikation Kommunikationsgrundrechte einschränkt, muss dem in Art. 5 Abs. 2 GG niedergelegten Erfordernis der „Allgemeinheit" genügen, das Maßnahmen gegen einzelne Meinungen ausschließt.[95] Ein solches Gesetz ist somit zulässig, wenn es Meinungsneutralität gegenüber den politischen Strömungen und Weltanschauungen wahrt.[96]

Abseits legislativer Maßnahmen kann die staatliche Öffentlichkeitsarbeit eine Rolle im staatlichen Umgang mit Desinformationen spielen. So wurden z. B. „vielfältige Maßnahmen durchgeführt", um die Öffentlichkeit hinsichtlich der Vorgehensweise künstlich gesteuerter Desinformationskampagnen zu sensibilisieren.[97] Die staatliche Öffentlichkeitsarbeit ist, wie das BVerfG speziell in Bezug auf die Regierung und gesetzgebende Körperschaften aus-

91 S. z. B. *Milker*, ZUM 2017, 216 (220); *Löber/Roßnagel*, MMR 2019, 493 (496); *Bull*, Der Staat (58) 2019, 57 (93).

92 *Libertus*, ZUM 2018, 20 (26).

93 *Kind u.a.* 2017, 36.

94 2019 haben soziale Medien die Online-Nachrichtenmagazine als Ressource für Online-Nachrichten überholt. Soziale Medien sind jedoch lediglich für 3 Prozent der Menschen in Deutschland die einzige Ressource und immerhin für 10 Prozent Hauptnachrichtenquelle, s. *Hölig/Hasebrink* 2019, 20f.; s. auch *Stark/Magin/Jürgens*, UFITA 2018, 103 (116 ff.).

95 BVerfGE 71, 206 (214) – Sitzblockaden 1.

96 S. BVerfGE 7, 198 (209) – Lüth; *Schemmer*, in: Epping/Hillgruber 2019, Art. 5 GG Rn. 99; *Hillgruber*, JZ 2016, 495 (496). Als einzige Ausnahme erkennt das BVerfG die Befürwortung der nationalsozialistischen Gewalt- und Willkürherrschaft, BVerfGE 124, 300 (321 ff.) – Wunsiedel.

97 BT-Drs. 19/2224, 6.

führt, „nicht nur zulässig, sondern auch notwendig, um den Grundkonsens im demokratischen Gemeinwesen lebendig zu erhalten. Darunter fällt namentlich die Darlegung und Erläuterung der Politik der Regierung hinsichtlich getroffener Maßnahmen und künftiger Vorhaben angesichts bestehender oder sich abzeichnender Probleme sowie die sachgerechte, objektiv gehaltene Information über den Bürger unmittelbar betreffende Fragen und wichtige Vorgänge auch außerhalb oder weit im Vorfeld der eigenen gestaltenden politischen Tätigkeit".[98]

Die Kompetenz für staatliche Informationstätigkeit ergibt sich aus der Aufgabenzuweisung selbst oder als Annex zu ihr.[99] In materieller Hinsicht ist das Gebot der Richtigkeit und Sachlichkeit zu beachten. Die Informationstätigkeit muss dabei einen teils schwierigen Balanceakt ausführen, soll sie doch die politische Willensbildung vom Volk zu den Staatsorganen fördern, auf eine lenkende Einflussnahme auf den Meinungsbildungsprozess aber verzichten und nicht zu einer undemokratischen Willensbildung von oben nach unten führen.[100] Informationsversorgung kann insbesondere geboten sein, wenn ein besonderes Risiko der Einseitigkeit der bereits vorhandenen Informationen besteht. Dies gilt auch bei kurzfristig auftretenden Krisen und Herausforderungen, um den Bürgern Orientierungshilfen zu geben. Diese Maßnahmen sollen nur unter der Prämisse erfolgen, dass sie den „Bürger zur eigenverantwortlichen Mitwirkung an der Problembewältigung befähigen".[101]

Vor allem kann und soll die Regierung durch ihre Öffentlichkeitsarbeit die Bürgerinnen und Bürger zur eigenverantwortlichen Mitwirkung an der Problembewältigung befähigen. Dementsprechend erwarten die Bürger für ihre persönliche Meinungsbildung und Orientierung von der Regierung Informationen, wenn diese andernfalls nicht verfügbar wären. Dies kann insbesondere Bereiche betreffen, in denen die Informationsversorgung der Bevölkerung auf interessengeleiteten, mit dem Risiko der Einseitigkeit verbundenen Informationen beruht und die gesellschaftlichen Kräfte nicht ausreichen, um ein hinreichendes Informationsgleichgewicht herzustellen.[102]

Davon zu unterscheiden sind spezifische Befugnisse der Gefahrenabwehr im Zusammenhang mit der Verbreitung von Desinformationen.[103] Hierzu

98 BVerfGE 138, 102 (114) – Schwesig m.w.N.; 63, 230 (243).
99 BVerfGE 105, 252 – Glykol; BVerwGE 159, 327; *Seckelmann* 2018, 26 m.w.N.
100 BVerfGE 20, 56 (99) – Parteienfinanzierung I; 44, 125 (139f.).
101 BVerfGE 105, 252 – Glykol.
102 BVerfGE 105, 252 – Glykol.
103 Vgl. BT-Drs. 19/2224, 12.

zählen etwa besondere nachrichtendienstliche Befugnisse, die „über die bloße Teilhabe staatlicher Funktionsträger an öffentlichen Auseinandersetzungen oder an der Schaffung einer hinreichenden Informationsgrundlage für eine eigenständige Entscheidungsbildung der Bürger“ hinausgehen.[104] Zu nennen sind außerdem die Polizei- und Ordnungsbehörden, die im Einzelfall eine wichtige Funktion einnehmen können, Falschinformationen im Rahmen ihrer Zuständigkeit zur Abwehr von Gefahren für die öffentliche Sicherheit und Ordnung zu korrigieren. Der professionelle Umgang von Behörden und Institutionen im digitalen Raum gilt als ein Schlüssel bei der Bekämpfung von Desinformation.[105]

C. Bekämpfung von Desinformation durch Betreiber von Social Networks

Desinformation und der Einsatz von Malicious Bots finden vor allem in Social Networks statt.[106] Diese haben daher eine herausragend wichtige Rolle, Desinformationen zu erkennen und zu bekämpfen. Niemand ist dazu so geeignet und in der Lage wie sie.

I. Rolle der Social Networks bei der (Des-)Informationsverbreitung

Die Betreiber haben, anders als Außenstehende, die Möglichkeit, auch koordinierte Desinformationskampagnen und technische Manipulationsformen wie Malicious Bots aufzuspüren. Zudem beeinflussen und kontrollieren die großen Social Networks faktisch auch wesentliche Teile der demokratischen Willensbildung, indem sie öffentliche Kommunikationsräume zur Verfügung stellen und gestalten. Sie bestimmen auf ihren Plattformen die Verwirklichungsbedingungen der Kommunikationsgrundrechte mit, indem sie beispielsweise Nachrichtenmeldungen individualisiert bereitstellen und Beiträge und Nutzerkonten auf Grundlage ihrer eigenen Regelwerke (AGB) löschen oder sperren. Dies wirft zum einen die Frage auf, in welchem Verhältnis privatautonomes Recht zu demokratischem Recht stehen sollte, wenn es um die öffentliche Kommunikation und die demokratische Willensbildung geht. Es provoziert zum anderen die Frage, ob die Informationssteuerung und

104 Vgl. BVerfGE 113, 63.
105 Näher dazu *Sängerlaub* 2018, 10 ff.
106 S. Kap. 1.C.

-filterung durch Algorithmen durch demokratische Rechtsnormen gelenkt werden muss.[107] Zum dritten sind damit Fragen zur Reichweite und zu Auswirkungen der mittelbaren Drittwirkung der Grundrechte und des Schutzauftrags des Staates für diese angesprochen, der er nur durch eine Verpflichtung der Plattformbetreiber genügen kann. Schließlich ist das Grundrecht der Meinungsfreiheit derjenigen angesprochen, gegen die Kommunikationsplattformen wegen der Verbreitung von Desinformationen vorgehen, indem sie etwa deren Beiträge löschen oder sperren.

II. Gesetzliche Lösch- und Sperrverpflichtungen

Wesentliche rechtliche Pflichten ergeben sich für Social Networks aus der sog. *Host-Providerhaftung*. Anbieter von Telemediendiensten, die fremde Informationen für einen Nutzenden speichern (sog. Host-Provider i.S.d. § 10 Telemediengesetz (TMG)), ohne sich diese zu eigen zu machen, müssen nach § 7 Abs. 2 Satz 1 TMG nicht proaktiv die Nutzerbeiträge auf Rechtsverstöße kontrollieren. Sie müssen erst unverzüglich tätig werden und rechtswidrige Inhalte entfernen, wenn sie Kenntnis von diesen, z.B. durch Nutzermeldung, erlangt haben.[108] Diese Regelung gilt jedoch nicht für eine Haftung des Providers als mittelbarer Störer.[109] So beinhaltet dessen Prüfungs- und Kontrollpflicht nach deutscher Rechtsprechung auch die Verpflichtung, nach erfolgtem Hinweis auf eine klare Rechtsverletzung weitere Verletzungen zu verhindern, soweit dies möglich und zumutbar ist.[110] Jene Provider sind daher zu einer proaktiven Suche nach weiteren identischen oder sinngleichen rechtswidrigen Inhalten, auf die sie nicht konkret hingewiesen worden sind, verpflichtet.[111]

Das *Netzwerkdurchsetzungsgesetz* (NetzDG) soll die Betreiber von Social Networks mit mehr als zwei Millionen registrierten Nutzenden in Deutschland dazu anhalten, ihren Pflichten zur Löschung und Sperrung von gemelde-

107 S. *Hoffmann-Riem*, AöR 2017, 1 (11); *ders*. 2017, 6 ff.
108 Ausführlich EuGH, GRUR 2012, 265 (267 ff.); *Holznagel* 2013; *Jandt*, in: Roßnagel 2013, § 10 TMG, Rn. 8 ff.
109 S. näher *Jandt*, in: Roßnagel 2013, § 10 TMG, Rn. 59 ff.
110 S. BGH, MMR 2015, 674 Rn. 49; *Gounalakis*, NJW 2013, 2321 (2323).
111 Zur Vereinbarkeit mit der E-Commerce-RL 2000/31/EG vom 8.6.2000 s. die Vorlagefragen des ÖOGH, ZUM 2018, 395 und die dazugehörige Entscheidung EuGH, Urt. v. 3.10.2019-C-18/18, ECLI:EU:C:2019:821 Rn. 33 ff. – Glawischnig-Piesczek/Facebook; LG Würzburg, MMR 2017, 347 (349).

ten strafbaren Falschnachrichten und Hassbeiträgen nachzukommen.[112] Kern des NetzDG ist es, dass jeder Betreiber ein Beschwerdemanagementsystem vorhalten muss, das es gewährleistet, dass strafbare Inhalte innerhalb einer bestimmten Frist gelöscht werden.[113] Offensichtlich strafbare Inhalte müssen sie innerhalb von 24 Stunden ab Meldung des Inhalts, schwieriger zu beurteilende strafbare Inhalte müssen sie innerhalb von sieben Tagen löschen. Die erfassten strafbaren Inhalte sind in § 1 Abs. 3 NetzDG aufgelistet. Über ihr Beschwerdemanagement zum Umgang mit Meldungen müssen die Anbieter gemäß § 2 NetzDG halbjährlich öffentlich Bericht erstatten.

Die verpflichtende Berichterstattung hat eine begrüßenswerte Transparenzsteigerung bewirkt. So verdeutlichen die Ausführungen der Social Networks in den Transparenzberichten, dass sie den eigenen Kriterien aus privater Regulierung Vorrang vor den staatlichen Vorgaben einräumen: Nach dem Eingang einer Meldung erfolgt eine zweistufige Prüfung.[114] Die Plattformen prüfen zuerst, ob die gemeldeten Inhalte gegen ihre eigenen Regelwerke (AGB) verstoßen (1. Stufe). Bejahen sie dies, löschen sie den Inhalt weltweit. Nur wenn dies nicht der Fall ist, prüfen sie die Inhalte entsprechend dem NetzDG am Maßstab des deutschen Strafrechts (2. Stufe). Diese werden nur für Nutzerinnen und Nutzer in Deutschland gesperrt. Die Anbieter halten somit an ihren bisherigen, etablierten Prüfverfahren fest und führen die spezifische NetzDG-Prüfung nachgelagert nur für die verbleibenden Inhalte durch.[115]

Den Angaben von YouTube ist zu entnehmen, dass ein ganz überwiegender Teil der gemeldeten und entfernten Inhalte bereits auf der 1. Stufe wegen eines Verstoßes gegen die AGB gelöscht wird.[116] Doch gerade in der Kategorie „Hassrede oder politischer Extremismus" ist der Anteil der NetzDG-Sperrungen bei YouTube höher als in den meisten anderen Kategorien. Hier waren die deutschen Strafgesetze bei etwa einem Viertel der Entfernungen (24,28 %) strenger als die Community-Richtlinien. Auch in der Kategorie „terroristische oder verfassungswidrige Inhalte" wurden etwas mehr als ein Drittel der Entfernungen (36,34 %) nach deutschem Strafrecht gesperrt und nicht auf Grundlage der Community-Richtlinien gelöscht.

Ein wesentlicher Kritikpunkt am NetzDG lautet, das Gesetz setze durch die Kombination von hoher Bußgeldandrohung und zeitlichem Entscheidungs-

112 Vgl. BT-Drs. 18/12356, 1 ff.
113 S. hierzu ausführlich *Roßnagel u.a.* 2018.
114 *Facebook* 2019, 6f.; textitTwitter 2019, 12; *YouTube* 2019, 1.
115 S. *Löber/Roßnagel*, MMR 2019, 71f.
116 *YouTube* 2019, 10.

druck unzulässige Anreize zum „Overblocking" von Nutzerbeiträgen und verstoße daher gegen die Meinungs- und Informationsfreiheit.[117] Jedoch gilt die Bußgeldandrohung nur für systemische, schwere Mängel im Beschwerdemanagement der Social Networks, nicht für einzelne Sperrungen.[118] Zudem lässt die vorgetragene Kritik, das NetzDG würde schon strukturell Anreize zur übermäßigen Entfernung von Inhalten geben, unberücksichtigt, dass in den Social Networks auch durchaus Tendenzen zum „Underblocking" bestehen können, denen das NetzDG entgegenwirkt.[119] Auch die Angaben in den Transparenzberichten und Lösch- und Sperrquoten von 23,38 % bei YouTube[120], 33,24 % bei Facebook[121] und nur 14,24 % im Falle von Twitter[122] bezogen auf die Meldungen können Overblocking nicht belegen.[123] Im Hinblick auf die vom Gesetzgeber mit dem NetzDG bezweckte Effektivierung der Durchsetzung des geltenden Rechts im Bereich der Hasskriminalität und strafbarer Falschnachrichten[124] ist davon auszugehen, dass diese jedenfalls gesteigert wurde, nicht zuletzt durch die Erhöhung der Mitarbeitenden sowie verbesserte Prüfverfahren der Social Networks.[125]

Soweit das NetzDG auch der Bekämpfung von Desinformationen dienen soll,[126] kann es erwartungsgemäß nur einen begrenzten Beitrag leisten.[127] Da viele Desinformationen nicht strafbar sind, ist das Gesetz von vornherein nicht geeignet, diese zu bekämpfen. Allerdings bewirkt es hinsichtlich der Entfernung von persönlichkeitsrechtsverletzenden Falschbehauptungen eine verbesserte Rechtsdurchsetzung und trägt damit zum Persönlichkeitsschutz bei. Dies gilt nicht nur wegen der durch die Nutzermeldung aktivierten und effektivierten Prüfpflicht der Social Networks, sondern auch aufgrund der mit Art. 2 NetzDG vollzogenen Ergänzung des § 14 TMG zur Datenherausgabe: Danach haben die Diensteanbieter nicht mehr nur den zuständigen Behörden

117 So u.a. *Guggenberger*, NJW 2017, 2577 (2581); *Gersdorf*, MMR 2017, 439 (446f.); *Müller-Franken*, AfP 2018, 1 (7f.); *Papier*, NJW 2017, 3025 (3030); *Steinbach*, JZ 2017, 653 (660).
118 S. *Roßnagel u.a.* 2018, 7.
119 S. *Lang*, AöR 2018, 220 (236); *Schiff*, MMR 2018, 366 (370); *Drexl*, ZUM 2017, 529 (533).
120 S. *YouTube* 2019, 3, 5.
121 S. *Facebook* 2019, 4, 10.
122 S. *Twitter* 2019, 12.
123 *Löber/Roßnagel*, MMR 2019, 71 (73).
124 BT-Drs. 18/12356, 1, 11.
125 *Löber/Roßnagel*, MMR 2019, 71 (74); s. auch *Deutscher Bundestag Parlamentsnachrichten* 2018.
126 BT-Drs. 18/12356, 1, 11.
127 S. *Holznagel*, MMR 2018, 18 (21).

zwecks Strafverfolgung und Gefahrenabwehr im Einzelfall Auskunft über die Bestandsdaten zu erteilen, sondern gemäß Abs. 3 und 4 unter Richtervorbehalt auch denjenigen, die zivilrechtliche Ansprüche wegen der Verletzung absolut geschützter Rechte aufgrund rechtswidriger Inhalte nach § 1 Abs. 3 NetzDG durchsetzen möchten.

Als problematisch wird erachtet, dass nach den Umständen des Einzelfalls aufwendigere Recherchen zur Verifizierung der (Un)wahrheit der Behauptung nötig sein können. Nach Einschätzung der Social Networks war es jedoch gar nicht oder nur in seltenen Ausnahmefällen erforderlich, die Stellungnahme des Uploaders des Inhalts gemäß § 3 Abs. 2 Nr. 3 lit. a NetzDG einzuholen. Ob darüber hinaus Faktenchecks stattgefunden haben, geht aus den Berichten nicht hervor. Weitere Recherchen werden vom NetzDG aber auch nicht gefordert.

III. Rechtliche Verantwortung der marktmächtigen Kommunikationsplattformen

Für die Bestimmung, inwieweit die Anbieter von Social Networks eine privatautonome Rechtsordnung errichten dürfen und inwieweit sie dabei Grundrechte der Nutzenden berücksichtigen müssen, ist ihre rechtliche Verantwortung und ihr Verhältnis zur demokratischen staatlichen Rechtsordnung zu klären.

Die grundrechtliche Freiheit, auf die sich die Unternehmen berufen können, wird durch die Grundrechte der Nutzenden und den Vorrang der demokratischen Rechtssetzung begrenzt. Die Grundrechte der Nutzenden gelten nicht nur gegen den Staat, sondern entfalten über die zivilrechtlichen Generalklauseln, z. B. § 241 Abs. 2 BGB, auch eine Drittwirkung auf das Privatrechtsverhältnis zwischen Anbieter und Nutzendem.[128] Die Bedeutung der Grundrechte – sowohl zum Schutz der Verletzten als auch zum Schutz der (Des-)Informationen Verbreitenden – wird derzeit angesichts der steigenden Macht der Plattformbetreiber über privat beherrschte öffentliche Kommunikationsräume neu vermessen. Die (verfassungs-)rechtliche Verpflichtung und Begrenzung der Inhalteregulierung, insbesondere für die Entfernung von Nutzerinhalten durch Social Networks, sind Gegenstand jüngerer Recht-

128 BVerfGE 7, 198 (205) – Lüth.

sprechung und vielfältiger Literaturbeiträge. Jedoch sind diese Fragen noch nicht durch höchstrichterliche Rechtsprechung abschließend geklärt.[129]

Die Gerichte befassen sich – zumeist im Kontext von einstweiligen Verfügungsverfahren wegen Löschungen aufgrund der Hassrede-Regelung von Facebook – damit, aus der wirkmächtigen Stellung großer Kommunikationsplattformen und den grundgesetzlichen Wertungen konkrete Schlussfolgerungen für die Reichweite der Befugnisse zur eigenen Rechtssetzung und -durchsetzung abzuleiten. Im Ausgangspunkt einig sind sich die Gerichte bei der Einordnung von Facebook, das als „möglicherweise sogar marktbeherrschende Plattform"[130] „einen öffentlichen Marktplatz für den Meinungs- und Informationsaustausch"[131] und „einen öffentlichen Kommunikationsraum"[132] zur Verfügung stellt. Unter Rückgriff auf das obiter dictum im Fraport-Urteil des Bundesverfassungsgerichts aus dem Jahr 2011, nach dem zum Schutz der Kommunikation „je nach Gewährleistungsinhalt und Fallgestaltung" „die mittelbare Grundrechtsbindung Privater einer Grundrechtsbindung des Staates (…) nahe oder auch gleich kommen" kann, wenn das private Unternehmen in wesentlichem Maß die Rahmenbedingungen öffentlicher Kommunikation übernimmt,[133] wird die Anwendung und Konsequenz dieser Rechtsprechung für wirkmächtige Social Networks im privatrechtlichen Verhältnis zu ihren Nutzerinnen und Nutzern diskutiert. Auch wird der Stadionverbot-Beschluss aufgegriffen,[134] in dem das Bundesverfassungsgericht 2018 für spezifische Konstellationen gleichheitsrechtliche Anforderungen aus Art. 3 Abs. 1 GG für das Verhältnis zwischen Privaten annimmt.[135] Damit entwickelte das Gericht die – traditionell auf Freiheitsrechte angewendete – Dogmatik und Reichweite der Drittwirkung im Hinblick auf die Anwendung von Gleichheitsrechten in Privatrechtsverhältnissen weiter und folgte dem in vorigen Entscheidungen vorgezeichneten Weg zur Stärkung der mittelbaren Grundrechtswirkung und Grundrechtsverpflichtung Privater in Sonderkonstellationen.

129 BVerfG, NVwZ 2019, 959 Rn. 15.
130 LG Frankfurt a.M., MMR 2018, 770 Rn. 43; BVerfG, NVwZ 2019, 959 Rn. 15 „erhebliche Marktmacht".
131 OLG München, MMR 2018, 760 Rn. 26 und NJW 2018, 3115 Rn. 26; LG Frankfurt, MMR 2018, 770 Rn. 43.
132 OLG Dresden, NJW 2018, 3111 Rn. 19.
133 BVerfGE 128, 226 (248 ff.) – Fraport; s. auch BVerfG, NJW 2015, 2485 (2486) – Bierdosenflashmob.
134 LG Frankfurt a.M., MMR 2018, 545 Rn. 11 ff.
135 BVerfG, NJW 2018, 1667 Rn. 41 – Stadionverbot; s. auch *Michl*, JZ 2018, 910; *Hellgardt*, JZ 2018, 901.

Zum Teil wird die Auffassung vertreten, ein Verstoß gegen die plattformeigenen Regelungen dürfe nicht zu einer Löschung führen, wenn der Beitrag die Grenzen zulässiger Meinungsäußerung einhalte.[136] Im Ergebnis wird damit eine unmittelbare Grundrechtsverpflichtung wirkmächtiger Kommunikationsplattformen gefordert. Andere Ansichten gehen in divergierenden Ausprägungen von einer weitergehenden Befugnis bei der Inhalteregulierung aus.[137] Der abstrakt generelle Ausschluss bestimmter Inhalte durch plattformeigene Regelwerke sei „als Ausübung der von Art. 2, 12, 14 GG geschützten Freiheiten der Anbieter ohne weiteres zulässig, und zwar gerade auch dann, wenn bestimmte Inhalte verboten werden sollen, die nach der Rechtsordnung legal sind".[138] So dürften in Ausübung des virtuellen Hausrechts auch Inhalte durch die Hassrede-Regelung von Facebook verboten werden.

Zu befürworten ist, dass wirkmächtige Anbieter von Social Networks – und damit großer öffentlicher Kommunikationsräume – einer strengen, intensivierten mittelbaren Drittwirkung in Bezug auf die Kommunikationsfreiheiten und den Persönlichkeitsschutz unterliegen. Sie müssen die Grundrechte ihrer Nutzenden weitestgehend achten. Sie unterliegen aber trotz ihrer besonderen rechtlichen Verantwortung keiner staatsgleichen Bindung. Ihnen darf nicht völlig versagt werden, in ihren Richtlinien eigene Akzente zu setzen. Die grundrechtliche Stellung der Anbieter von Social Networks muss in der Abwägung der Grundrechtspositionen hinreichend berücksichtigt werden. Daher kann ein berechtigtes Interesse an der weitergehenden Regulierung insbesondere bestehen, wenn das Unternehmen durch die Entfernung des Beitrags einen eigenen politischen Standpunkt vertreten will oder erhebliche geschäftliche Interessen betroffen sind.[139] Dabei ist auch zu berücksichtigen, dass Hassbeiträge, denen es aus Sicht der Anbieter entgegenzuwirken

136 OLG München, MMR 2018, 760 Rn. 26 und NJW 2018, 3115; KG, BeckRS 2019, 9590 Rn. 17 für YouTube; LG Frankfurt a.M., MMR 2018, 545 Rn. 14; LG Karlsruhe, BeckRS 2018, 20324 Rn. 12; *Pille* 2016, 360f. Nach OLG München, MMR 2019, 469 Ls. 3, soll dies nicht gelten für Beiträge in einem (Unter-)Forum einer Social-Media-Plattform „mit eindeutig erkennbarer begrenzter Zweckbestimmung", die den dort diskutierten Themen nicht zuzuordnen sind („Themaverfehlung").

137 OLG Dresden, NJW 2018, 3111; OLG Karlsruhe, NJW 2018, 3110; OLG Stuttgart, MMR 2019, 110 (112); *Elsaß/Labusga/Tichy*, CR 2017, 234 (235 ff.); *Beurskens*, NJW 2018, 3418 (3419); *Holznagel*, CR 2018, 369 (371f.); *Spindler*, CR 2019, 238 (243ff.).

138 OLG Dresden, NJW 2018, 3111 Rn. 15; i.E. ebenso OLG Karlsruhe, NJW 2018, 3110 (3111).

139 Eingehend *Raue*, JZ 2018, 961 (967 ff.); s. auch *Lüdemann*, MMR 2019, 279 (281f.).

gilt, negative Auswirkungen auf den dort stattfindenden Meinungsaustausch, die Ermöglichung sachbezogener Diskussionen und die freie Rede für alle Nutzenden haben und letztlich zur Infragestellung des Geschäftsmodells der Kommunikationsplattform führen können.[140] Schließlich ist die Meinungsfreiheit von anderen Nutzenden auch betroffen, wenn durch Hassrede eine Diskussion nachhaltig negativ beeinflusst wird, sodass sie eingeschüchtert werden und von einer (weiteren) Beteiligung absehen. Gleichheitsrechtliche Anforderungen sind vom Social Network zu beachten, wenn eine marktbeherrschende Stellung besteht und der Betroffene stark auf die Mitgliedschaft im Social Network angewiesen ist.[141] An den sachlichen Grund sollten jedoch keine umfassenden Anforderungen gestellt werden, um die Unterscheidung von grundrechtsverpflichtetem Staat und privatem Akteur nicht völlig zu ignorieren.

Für die Anforderungen an einen rechtmäßigen Umgang mit Desinformationen bedeutet dies, dass die Netzwerkbetreiber zum einen (weiterhin) eigene Regelungen mit Bezug zu Desinformationen treffen dürfen. Diese müssen aber einer AGB-Kontrolle standhalten, die auch die gesteigerte mittelbare Drittwirkung berücksichtigt. Ihr Handlungsspielraum ist umso geringer, je stärker die Inhalte von den Kommunikationsgrundrechten erfasst sind. Beispielsweise dürfen Satire-Inhalte, die als solche erkennbar sind, auch dann nicht entfernt werden, wenn sie falsche Informationen enthalten.[142] Vielfach muss der Betreiber eine schwierige Abwägung zwischen den konkurrierenden Grundrechten seiner Nutzenden – einerseits auf Schutz ihrer Persönlichkeit und auf die Integrität ihrer Kommunikation und andererseits auf das Recht zur Meinungsäußerung der informierenden Nutzenden – sowie seiner eigenen Grundrechte durchführen. Dem Rationalisierungs- und Standardisierungsinteresse der Betreiber sind somit Grenzen gesetzt.[143] Der Überprüfungsaufwand wird für sie zudem nicht bereits dadurch unzumutbar, dass sie zusätzliches Personal für die Kontrolle einsetzen müssen, sondern erst, wenn durch ihn das Geschäftsmodell in Frage gestellt würde.[144]

140 S. OLG Dresden, NJW 2018, 3111 Rn. 20; LG Frankfurt a.M., BeckRS 2018, 21919 Rn. 24.
141 Vgl. BVerfG, NVwZ 2019, 959 Rn. 15.
142 S. 5.A.II.; LG Nürnberg-Fürth, BeckRS 2019, 12259 Rn. 26 ff.
143 Zur Zulässigkeit von Verdachtsklauseln, d.h. die Befugnis, bei objektivierbaren Verdacht im Hinblick auf die Unzulässigkeit von Inhalten zu sperren oder zu löschen, s. *Holznagel*, CR 2018, 369 (373f. m.w.N.).
144 S. LG Würzburg, MMR 2017, 347 (349).

IV. Automatisierte Prozesse und Entscheidungen

Aus der Erforderlichkeit einer Einzelfallprüfung ergeben sich deutliche Begrenzungen für die Möglichkeit des rechtmäßigen Einsatzes von automatisierten Prozessen im Kontext von Desinformation. Das automatisierte Auffinden von Desinformation kann als Hilfswerkzeug dienen. Die Entscheidung über das weitere Vorgehen müssen regelmäßig geschulte Mitarbeiterinnen und Mitarbeiter übernehmen.

Der Einsatz von technischen Meldesystemen spielt in Social Networks eine wichtige Rolle und unterliegt einer stetigen Weiterentwicklung. Falschmeldungen identifiziert Facebook mit einer Kombination aus technischen Verfahren, vornehmlich maschinelles Lernen, und menschlicher Verifizierung.[145] Inwieweit Facebook in die Entscheidung darüber, wie mit automatisiert identifizierten Hassreden oder Falschmeldungen verfahren wird – ob also ein Hassredebeitrag entfernt,[146] oder die Sichtbarkeit von Falschmeldungen im News Feed reduziert wird[147] Menschen einbindet, bleibt angesichts vager Angaben jedoch unklar. Zum Teil unterstützen externe Faktenprüfer Facebook bei der Identifizierung irreführender Inhalte. Maschinelles Lernen setzt das Unternehmen etwa ein, um Duplikate bereits von Faktenprüfern widerlegter Beiträge aufzuspüren. Andererseits weist es auch darauf hin, dass der manuellen Überprüfung von Beiträgen angesichts der täglich mehr als eine Milliarde veröffentlichten Inhalte auf der Plattform Grenzen gesetzt sind und es daran arbeitet, automatisierte Systeme zu verbessern.[148]

Begrenzungen des Handlungsspielraums, automatisiert Entscheidungen zu treffen, ergeben sich auch aus datenschutzrechtlichen Erwägungen. So regelt Art. 22 Abs. 1 DSGVO, dass Personen das Recht haben, nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Die Vorschrift statuiert ein grundsätzliches, von individueller Geltendmachung unabhängiges Verbot ausschließlich automatisierter Entscheidungen im Einzelfall, damit der

145 *Lyons*, „Verstärkung unserer Anstrengungen gegen Falschmeldungen", Facebook Newsroom v. 21.6.2018, https://de.newsroom.fb.com/news/2018/06/verstaerkung-unserer-anstrengungen-gegen-falschmeldungen/.

146 Gemeinschaftsstandards von *Facebook*, III. Ziff. 11 „Hassrede".

147 Gemeinschaftsstandards von *Facebook*, IV. Ziff. 18 „Falschmeldungen".

148 „Facebook weitet Faktenprüferprogramm in Deutschland aus – Deutsche Presse-Agentur wird neuer Partner", Facebook Newsroom v. 18.3.2019, https://de.newsroom.fb.com/news/2019/03/dpa_faktenpruefer/.

Einzelne nicht zum bloßen Objekt rein maschineller Entscheidung wird.[149] Das Verbot dient außerdem dem Schutz vor diskriminierenden Entscheidungen von vermeintlich objektiven Datenverarbeitungsprogrammen sowie der Förderung von Transparenz und Fairness bei der Entscheidungsfindung.[150] Deutliche Beschränkungen des Anwendungsbereichs des Verbots bestehen aus mehreren Gründen.

Der Anwendungsbereich ist dem Wortlaut nach nur eröffnet für automatisierte Entscheidungen, die eine rechtliche Wirkung (Alt. 1) oder ähnliche erhebliche Beeinträchtigung (Alt. 2) für die betroffene Person erzeugen. Die Tatbestandsvariante der rechtlichen Wirkung ist – auch um der zweiten Variante der „erheblichen Beeinträchtigung" noch Raum zu geben – eng auszulegen und auf solche Entscheidungen begrenzt, die eine Rechtsfolge auslösen, namentlich nur bei einseitigem hoheitlichem oder einseitigem rechtsgeschäftlichem Handeln.[151]

Sollte ein Nutzerkonto entfernt und dadurch der Nutzungsvertrag durch das Social Network gekündigt werden, zieht die automatisierte Entscheidung eine Rechtsfolge nach sich, sodass eine rechtliche Wirkung im Sinne der Vorschrift vorliegen dürfte. Auch Sperrungen und Löschungen von Beiträgen sind gestaltende Akte, von denen die Ersteller der Beiträge individuell getroffen werden. Soweit Löschungen und Sperrungen von Beiträgen eine rechtliche Wirkung i.S.d. Vorschrift nicht zukommen sollte, sind sie unter die Tatbestandsalternative der erheblichen Beeinträchtigung in ähnlicher Weise zu subsumieren. Löschungen und Sperrungen sind in der Regel Einschränkungen der Meinungsfreiheit der Betroffenen. Sie stellen Diskriminierungen der betroffenen Personen gegenüber denjenigen dar, die die Kommunikationsplattform weiterhin nutzen können. Denkbar ist auch eine erhebliche Störung der persönlichen Entfaltung. Hier könnten aber Ausnahmen geboten sein, etwa bei (besonders schweren) Verstößen gegen geltendes Recht oder Nutzungsbedingungen, sofern diese hinreichend sicher von dem automatisierten Prüfsystem erkannt werden können. Es kann differenziert werden zwischen Beiträgen, die stärkeren grundrechtlichen Schutz genießen, wie von Art. 5 GG geschützte Meinungsäußerungen, und solchen, die schon gar nicht grundrechtlich geschützt sind oder denen nur ein schwacher Schutz

149 *Abel*, ZD 2018, 304 (305); *Martini/Nink*, NVwZ 2017, 681 (681); *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 16 m.w.N.; ähnlich *Schulz*, in: Gola 2018, Art. 22 DSGVO Rn. 5.

150 *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 3.

151 *Lewinski*, in: Wolff/Brink 2019, Art. 22 DSGVO Rn. 28f.; *Schulz*, in: Gola 2018, Art. 22 DSGVO Rn. 22.

etwa über die allgemeine Handlungsfreiheit zukommt. So wären beispielsweise die Urheber strafbarer Beiträge nicht erheblich beeinträchtigt i.S.d. Vorschrift, ebenso wenig die Verwender von Social Bots, die massenhaft Beiträge generieren, sowie die Ersteller von Spam-Accounts.

Im Hinblick auf den Entscheidungsvorgang schließt Art. 22 Abs. 1 DSGVO nicht aus, dass überhaupt automatisierte Verfahren zum Einsatz kommen: Beispielsweise kann die Entscheidung im automatisierten Verfahren vorbereitet und ein automatisiert erzeugter Entscheidungsvorschlag von einem Menschen anhand weiterer Kriterien einer abschließenden Beurteilung unterzogen und Grundlage einer eigenen Entscheidung werden.[152] Somit gilt das Verbot nicht für Systeme, mit denen Entscheidungen über Maßnahmen zu Desinformationen nur vorbereitet oder empfohlen werden und ein Mensch die finale Entscheidung trifft, auch entgegen der automatisiert empfohlenen Entscheidung (Decision Support Systems).[153] Dies erscheint nicht zuletzt vor dem Hintergrund der Masse von Beiträgen in Social Networks sachgerecht. Wenn also menschliche Faktenprüfer die automatisiert aufgefundenen Informationen bewerten und auf Grundlage ihrer Entscheidung die Sichtbarkeit der für falsch befundenen Nachrichten eingeschränkt wird, ist dies mit Art. 22 Abs. 1 DSGVO vereinbar. Demgegenüber liegt aber in Anbetracht des Schutzzwecks der Norm auch dann eine automatisierte Entscheidung vor, wenn ein Mensch eine automatisierte Vorgabe lediglich bestätigt oder übernimmt, ohne eigene Erwägungen anzustellen, oder wenn etwa bloß Stichprobenkontrollen durchgeführt werden.[154]

Automatisierte Entscheidungssysteme können auch zum Einsatz kommen, wenn die Betreiber das Social Network automatisiert nach bestimmten rechtswidrigen Informationen durchsuchen, die wortgleich mit zuvor manuell als rechtswidrig erkannten Informationen sind,[155] und die auf diese Weise gefundenen Beiträge automatisiert entfernt werden. Hier wird das Verbot nicht greifen, da eine erhebliche Beeinträchtigung zu verneinen ist, wenn Nutzende mehrmals identische rechtswidrige Inhalte erstellen und auch Zumutbarkeitserwägungen auf Seiten des Netzwerkbetreibers zu berücksichtigen sind.

Zu beachten sind auch die in Art. 22 Abs. 2 lit. a bis c DSGVO vorgesehenen Ausnahmen vom Verbot automatisierter Entscheidungen. Eine

152 *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 28 m.w.N.
153 S. *Dreyer/Schulz* 2018, 19.
154 *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 26f.
155 S. dazu die Ausführungen unter 5.C.II.

ausdrückliche Einwilligung der betroffenen Person (lit. c)[156] dürfte für die hier interessierenden Fälle automatisierter Entscheidungen durch Social Networks nicht vorliegen.[157] Das Verbot findet zudem keine Anwendung, wenn die automatisierte Entscheidung für den Abschluss oder die Erfüllung eines Vertrags zwischen der betroffenen Person und dem Verantwortlichen erforderlich ist (lit. a). Die Vorschrift soll den praktischen Anforderungen von Massengeschäften Rechnung tragen, wobei die automatisierte Entscheidung nicht Vertragsgegenstand sein muss.[158] Die Erforderlichkeit ist zu bejahen, wenn die automatisierte Einzelentscheidung ein geeignetes Mittel zur Erreichung des Vertragszwecks ist und keine datenschutzrechtlich milderen, gleich wirksamen Mittel ersichtlich sind.[159] Zudem ist die Erforderlichkeit dann gegeben, wenn die automatisierte Entscheidungsfindung zur Erfüllung einer gesetzlichen Verpflichtung erfolgt.[160] Soweit man bei automatisierten Entscheidungen über massenhaft eingesetzte Social Bots, Spam-Accounts und Co. nicht bereits die Tatbestandsmäßigkeit nach Abs. 1 verneint, könnte die Ausnahme gemäß lit. a greifen, da diese geeignet sind, die Kommunikation auf den Plattformen erheblich zu stören, zu verfälschen und klare, schwerwiegende Verstöße gegen die AGB der Netzwerkbetreiber darstellen und angesichts der Masse der zu ergreifenden Maßnahmen sowie teilweise technisch fortgeschrittener Manipulationsformen nur mittels automatisierter Werkzeuge zügig Detektionen und Entscheidungen durchgeführt werden können. Die AGB beinhalten auch eine Selbstbindung des Netzwerkbetreibers,[161] die dort benannten Inhalte und Verhaltensweisen auf der Kommunikationsplattform nicht zu dulden.

Ersichtlich wird somit, dass das Verbot automatisierter Entscheidungen gemäß Art. 22 DSGVO in der Anwendung weniger streng ist, als es viel-

156 Diese Ausnahme wandelt faktisch das grundsätzliche Verbot nach Abs. 1 in ein Verbot mit Einwilligungsvorbehalt und könnte in der Praxis künftig weiter an Bedeutung gewinnen, um den Handlungsspielraum für automatisierte Entscheidungen zu erweitern, s. *Dreyer/Schulz* 2018, 22; *Enders*, JA 2018, 721 (725).

157 In den Nutzungsbedingungen von Facebook heißt es etwa lediglich: „Außerdem entwickeln wir automatisierte Systeme zur Verbesserung unserer Möglichkeit, missbräuchliche und gefährliche Aktivitäten, die unserer Gemeinschaft und der Integrität unserer Produkte schaden könnten, zu ermitteln und zu entfernen." Nutzungsbedingungen von *Facebook*, Stand Juli 2019, Ziff. 1, Unterpunkt 6.

158 *Schulz*, in: Gola 2018, Art. 22 DSGVO Rn. 30.

159 *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 42 m.w.N.

160 *Scholz*, in: Simitis/Hornung/Spiecker gen. Döhmann 2019, Art. 22 DSGVO Rn. 43.

161 *Beurskens*, NJW 2018, 3418 (3420).

leicht zunächst den Eindruck erweckt. Umso wichtiger sind die aufgezeigten Grenzen für automatisierte Entscheidungen, die sich insbesondere aus der Achtung der Meinungsfreiheit der Nutzenden ergeben.

V. Selbstregulierung

Im Frühjahr 2018 forderte die Europäische Kommission die Online-Plattformen auf, verstärkt Maßnahmen gegen Desinformation zu ergreifen, insbesondere einen „ehrgeizigen Verhaltenskodex" zu entwickeln.[162] Daraufhin beschlossen die großen Online-Plattformen und die Werbeindustrie im Herbst 2018 den „EU-Verhaltenskodex zur Bekämpfung von Desinformation". In diesem verpflichteten sie sich selbst, klare Strategien und Regelwerke gegen Desinformationen und Social Bots auf ihren Plattformen umzusetzen. Im Dezember 2018 forderte die Europäische Kommission die Social Networks im „Aktionsplan gegen Desinformation" dringender auf, unverzüglich Gegenmaßnahmen gegen Desinformation durchzuführen und automatische Bots zu identifizieren und entsprechend zu kennzeichnen. Sie kündigte weitere Maßnahmen, auch regulatorischer Art an, sollte die Umsetzung oder Wirkung des Kodex sich als unzureichend erweisen.[163]

Der Verhaltenskodex ist kein verbindliches Vertragswerk mit Sanktionsmöglichkeit, sondern enthält Selbstverpflichtungen der Online-Plattformen und Werbeindustrie (darunter auch Facebook, Twitter, Google). Die in dem Kodex enthaltenen Regelungen sind sehr vage und beziehen sich im Wesentlichen auf die bereits zuvor schon ausgeführten Maßnahmen der Plattformen. Ihnen fehlen ein gemeinsamer Ansatz, klare und bedeutsame Verpflichtungen sowie messbare Zielvorgaben. Die Unterzeichner haben sich selbst einen möglichst großen Spielraum eingeräumt. Diese Defizite sind auch darauf zurückzuführen, dass die Unterzeichner sich durchaus in einigen Punkten erheblich voneinander unterscheiden und es sich nicht nur um Anbieter von Social Networks handelt.

Der Kodex sieht eigentlich eine jährliche Berichtspflicht vor.[164] Im Kontext der EU-Wahlen im Mai 2019 haben die Betreiber monatliche Berichte vorgelegt, um Fortschritte zu zeigen, z. B. bei der Herstellung von Transparenz bei politischer Werbung. Die monatlichen Berichte haben die Transparenz

162 *Europäische Kommission*, „Bekämpfung von Desinformation im Internet: ein europäisches Konzept", Mitteilung vom 26.4.2018, COM(2018) 236 final, Ziff. 3.1.1.
163 *Europäische Kommission* 2018, 11.
164 EU-Verhaltenskodex zur Bekämpfung von Desinformation, 2018, 8, III. Nr. 16.

gesteigert. Allerdings beziehen sich viele Angaben gar nicht konkret auf die EU und es bleibt unklar, welche Auswirkungen Desinformationskampagnen auf sie haben. Für Social Bots haben die Social Networks, soweit ersichtlich, bislang keine allgemeinen Kennzeichnungssysteme eingeführt. Angaben zu Löschungen von Bots bleiben vage – sie beziehen sich in der Regel nicht konkret auf Deutschland und zumeist allgemein auf Fake-Accounts und Spam-Accounts, sodass unklar bleibt, inwieweit auch Social Bots erfasst sind. Bisher sind auch keine erheblichen Fortschritte im Hinblick auf die Herstellung eines datenschutzgerechten Zugangs für Forschungstätigkeiten[165] festzustellen.

D. Rechtspolitik

Da bisher weder die Maßnahmen der Europäischen Union noch der Mitgliedstaaten noch der Betreiber von Kommunikationsplattformen einen wesentlichen Fortschritt gebracht haben, Desinformationen und Verzerrungen der öffentlichen Diskussion rechtzeitig zu erkennen und zu bekämpfen, stellt sich die Frage, welche rechtspolitischen Strategien und Maßnahmen zusätzlich zu verfolgen sind.

Zwar fällt das Bekämpfen von offenkundigen Falschbehauptungen in die Eigenverantwortung der Bürgerinnen und Bürger und ist Sache des öffentlichen und politischen Meinungskampfes. Soweit jedoch Desinformationen und Verzerrungen des öffentlichen Diskurses die Möglichkeiten der Gegenwehr durch Bürger überschreiten und die freie Diskussion über öffentliche Themen gefährden, besteht eine verfassungsrechtliche Schutzpflicht, rechtliche Gegenmaßnahmen zum Schutz überragender Verfassungsgüter einzusetzen. Eine Befugnis für Vorsorgemaßnahmen, um die Entstehung solcher Gefährdungen zu verhindern, ist politisch sinnvoll und bei ausreichenden Anhaltspunkten für solche Entwicklungen zulässig. In der Umsetzung sind dann insbesondere der Verhältnismäßigkeitsgrundsatz sowie das Gebot schonender Mittelauswahl zu berücksichtigen.[166] Zu prüfen ist immer, ob auch geringere Grundrechtseingriffe wie Beobachtung, Erforschung und ähnliche Maßnahmen für die Bekämpfung der Risiken genügen.[167]

---

165 S. zum Datenschutz in der Forschung Roßnagel, ZD 2019, 157 und Kap. 4.B.II.
166 *Di Fabio*, in: Maunz/Dürig 2019, Art. 2 Abs. 2 Satz 1 GG, Rn. 90.
167 *Müller-Terpitz*, in: Isensee/Kirchhof 2009, § 147 Rn. 73.

I. Strafrechtliche Ergänzungen?

Viele Fälle von Desinformation sind nicht vom geltenden Strafrecht erfasst, da sie sich nicht auf individualisierbare Personen oder bestimmte geschützte Gruppen beziehen. Es könnte daher erwogen werden, den strafrechtlichen Schutz zu erweitern. Entsprechende Vorschläge, wie die Ergänzung von § 130 StGB um einen Tatbestand der Desinformation, die den öffentlichen Frieden und die öffentliche Rechtssicherheit stört,[168] sind mit großer Vorsicht zu betrachten. Fraglich ist insbesondere, welche Fälle hiervon tatsächlich erfasst sein sollten. Da die Hürde für ein strafbewehrtes Verhalten entsprechend hoch anzusetzen wäre, auch um den öffentlichen Diskurs, den Kampf der Meinungen, zu schützen, stünde zu befürchten, dass ein solcher Tatbestand kaum zur Anwendung kommen würde.[169]

Angesichts der Risiken für den Prozess der politischen Willensbildung könnte erwogen werden, die Integrität politischer Wahlen zu schützen. Beispielsweise könnten die §§ 108 ff. StGB um eine Regelung ergänzt werden, die Desinformation unter Strafe stellt, die geeignet ist, den Wählerwillen zu beeinflussen.[170] Solche Ansätze finden sich auch in ausländischen Rechtsordnungen. So hat der französische Gesetzgeber 2018 mit dem „Gesetz gegen Informationsmanipulation“[171] eine neue Regelungsgrundlage geschaffen, um die Verbreitung von Falschinformationen zur Wahlmanipulation stärker zu bekämpfen. Sie tritt neben existierendes Presse- und Wahlrecht, das bereits in gewissem Umfang Falschnachrichten verbietet, die mit der Absicht verbreitet werden, den öffentlichen Diskurs zu beeinflussen oder Wahlergebnisse zu verfälschen. Im Zentrum der Novelle stehen nicht Desinformationen per se, sondern solche, die in einem Zeitraum von drei Monaten vor einer Wahl im Rahmen von öffentlichen Online-Kommunikationsdiensten vorsätzlich, künstlich oder automatisiert verbreitet werden und die die Aufrichtigkeit der Wahl beeinträchtigen könnten. Im Falle eines Verstoßes können auf richterli-

168 *Mafi-Gudarzi*, ZRP 2019, 65 (67f.).

169 S. z. B. den inzwischen außer Kraft getretenen § 276 StGB (Österreich), der die Verbreitung falscher, beunruhigender Gerüchte unter Strafe stellte und für den die Kriminalstatistik seit 20 Jahren keine Verurteilung aufwies, s. *Schmid*, „Abgeschafft: Österreich hatte bis 2016 Gesetz gegen Fake-News“, Der Standard vom 20.12.2016, https://derstandard.at/2000049437140/Abgeschafft-Oesterreich-hatte-bis-2016-Gesetz-gegen-Fake-News.

170 *Mafi-Gudarzi*, ZRP 2019, 65 (68).

171 S. loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l’information.

che Anordnung alle angemessenen und notwendigen Maßnahmen getroffen werden, die unter Beachtung der Verhältnismäßigkeit geeignet erscheinen, eine Verbreitung der Falschinformation zu verhindern, d.h. insbesondere auch deren Löschung.[172]

Dennoch ist festzuhalten, dass der Anwendungsbereich solcher Regelungen relativ gering sein dürfte, insbesondere da die Pönalisierung von mit Werturteilen vermengten falschen Tatsachenbehauptungen wegen des Schutzes der Meinungsfreiheit nicht einzubeziehen wäre und die Beeinflussung des Willens der Wählerinnen und Wähler für sich genommen ein legitimes Anliegen verschiedener Akteure darstellt. Symbolische Gesetzgebung sollte vermieden werden. Die Androhung der Strafe muss bei Begehung der Straftat durchsetzbar sein, sodass ersichtlich wird, dass derjenige, der sich darüber hinwegsetzt, auch konsequent zur Rechenschaft gezogen wird. Die Ankündigung oder Androhung belastender Rechtsfolgen bildet zwar auch ein Motiv, unrechtes Verhalten zu unterlassen, wird alleine aber nicht zu einer effektiven Verhinderung führen.[173] Strafbewehrt sollte nur ein erheblich sozialschädliches Verhalten sein, das in entsprechender Ausprägung geeignet ist, mit elementaren Grundanforderungen des Gemeinschaftslebens zu brechen und somit besonders gravierende Begehungsformen tatbestandlichen Qualifikationen zugeordnet werden können.[174] Dies könnte nur für besonders verwerfliche oder gefährliche falsche Tatsachenbehauptungen als ultima ratio in Betracht kommen. Im Hinblick auf strafbare Online-Desinformationen und Internetkriminalität im Allgemeinen sollten Überlegungen zu neuen Straftatbeständen insbesondere nicht davon ablenken, dass eine wesentliche – wenn nicht die wichtigste – Herausforderung darin besteht, bereits geltendes Recht effektiv durchzusetzen.

II. Regulierung der Social Networks und Medienintermediäre

Zum Schutz der individuellen Freiheiten und zur „Sicherung der Funktionsfähigkeit der Kommunikationsordnung insgesamt“[175] sind staatliche Re-

172 Näher *Heldt*, „Von der Schwierigkeit, Fake News zu regulieren: Frankreichs Gesetzgebung gegen die Verbreitung von Falschnachrichten im Wahlkampf“, bpb.de vom 2.5.2019.

173 *Frisch*, NStZ 2016, 16 (17f.); allgemein zur Kritik an relativen Straftheorien s. *Schmitz-Remberg* 2014, 39 ff.

174 BVerfG, NJW 1998, 443 (443); *Frisch*, NStZ 2016, 16 (21).

175 *Hoffmann-Riem*, JZ 2014, 53 (56).

gelungen umso wichtiger, je stärker die Stellung eines Social Network und seine Bedeutung als öffentliches Forum für die Kommunikation und Meinungsbildung ist. Insofern sind die Regelungen des NetzDG grundsätzlich zu unterstützen.

Das *NetzDG* sollte jedoch entsprechend der folgenden Vorschläge verbessert werden: Es sollte sichergestellt sein, dass Social Networks, wie von § 3 Abs. 1 Satz 2 NetzDG gefordert, ein leicht erkennbares und unmittelbar erreichbares Verfahren für NetzDG-Meldungen bereithalten. Während z. B. YouTube und Twitter den „Melde-Button" für Meldungen nach dem NetzDG direkt bei der Nachricht angebracht haben, hat Facebook ihn so „versteckt", dass er nur schwer zu finden ist. Dementsprechend verzeichnete YouTube im ersten Halbjahr 2019 304.425 beanstandete Inhalte und Twitter 499.346 eingegangene Beschwerden, Facebook jedoch nur 674 Beschwerden mit 1.050 beanstandeten Inhalten.[176] Da dadurch in der Öffentlichkeit über das Ausmaß rechtswidriger Inhalte und die Art und Weise, wie Facebook mit ihnen umgeht, ein verzerrtes Bild entsteht, hat das Bundesamt für Justiz folgerichtig einen Bußgeldbescheid i.H.v. 2 Mio. Euro wegen Verstoßes gegen mehrere Regelungen des NetzDG erlassen.[177] Zu erwägen wäre, ob das NetzDG zusätzlich funktional festlegen sollte, dass die Möglichkeit einer Meldung direkt neben dem Inhalt zu finden sein muss, um Missverständnisse darüber, wann die genannten Anforderungen an die Meldemöglichkeit erfüllt sind, zu vermeiden.[178]

Das NetzDG berücksichtigt bislang nicht die Problematik der Löschung rechtskonformer Inhalte und sollte daher eine Regelung zu einem Put-Back-Verfahren enthalten, nach dem der Anbieter unrechtmäßig entfernte Inhalte wiederherstellen muss.[179] Über die Umsetzung dieser Regelung sollte der Anbieter ebenfalls berichten müssen.[180]

In den Berichten sollten die Plattformen auch die Anzahl der Inhalte, die sie auf Grundlage der plattformeigenen Regeln gelöscht haben, angeben. Die Frage des Over- und Underblocking kann nur dann überprüft und gesellschaftlich diskutiert werden, wenn die Plattformen auch anonymisier-

176 S. *YouTube* 2019, 3; *Facebook* 2019, 4; *Twitter* 2019, 12.

177 *Bundesamt für Justiz*, „Bundesamt für Justiz erlässt Bußgeldbescheid gegen Facebook", Pressemitteilung vom 2.7.2019 (in Bezug auf die Angaben von Facebook im Transparenzbericht über das 1. Hj. 2018).

178 S. hierzu *Löber/Roßnagel*, MMR 2019, 71 (72 ff.).

179 S. z. B. *Roßnagel u.a.* 2018, 12.

180 S. näher *Peukert*, MMR 2018, 572 (573 ff.).

te Einzelfälle mit Begründung der Entscheidung veröffentlichen.[181] Jedes Beschwerdeverfahren ist ohnehin nach § 3 Abs. 3 NetzDG zu dokumentieren.

Schließlich sollte das NetzDG eine Kooperation der Social Networks mit den Strafverfolgungsbehörden vorsehen. Diese könnte in einer Pflicht bestehen, gemeldete Inhalte, die von § 1 Abs. 3 NetzDG erfasste Straftaten darstellen, die keine Antragsdelikte sind, an die Strafverfolgungsbehörden weiterzuleiten. Bei 11.682 Sperrungen infolge des Meldegrundes Volksverhetzung allein in einem halben Jahr bei Twitter[182] muss sichergestellt werden, dass eine wirksame Strafverfolgung möglich ist.[183] Hierfür ist auch eine personelle Aufstockung der Strafverfolgungsbehörden unerlässlich.[184] Für die Bekämpfung von Hasskriminalität und strafbaren Falschnachrichten sind über das NetzDG hinaus weitere Maßnahmen erforderlich, die vor allem die faktischen Voraussetzungen der Strafverfolgung und des Rechtsschutzes der Nutzenden betreffen.[185]

Zu berücksichtigen sind auch Regulierungsansätze, die nicht spezifisch gegen Desinformation gerichtet sind, sondern sich mittelbar positiv auf deren Bekämpfung auswirken dürften, z. B. zur Meinungs- und Medienvielfalt. Im Rundfunk wird Meinungsvielfalt durch gesetzliche Vorgaben gewährleistet (§§ 11 Abs. 1 und 3, 25 Abs. 1 und 2 RStV).[186] Nunmehr schlägt der *Entwurf des Medienstaatsvertrags* (MStV-E) in § 53e ein allgemeines Diskriminierungsverbot vor, das zur Sicherung der Meinungsvielfalt die Behinderung oder Ungleichbehandlung journalistisch-redaktionell gestalteter Angebote durch Medienintermediäre untersagt.[187] Grundsätzlich erscheint es sinnvoll,[188] eine vielfaltssichernde Generalklausel zu etablieren, die an die Meinungsbildungsrelevanz des Medienintermediärs anknüpft und hiermit „ein System von abgestuften Verpflichtungen betreffend Auswahl, Reihung und Präsentation von Inhalten“ verbindet.[189]

---

181 *Eifert*, NJW 2017, 1450 (1453) fordert die Angabe der wesentlichen Umstände des Sachverhalts.
182 *Twitter* 2019, 16.
183 *Liesching*, MMR 2018, 26 (30).
184 S. BT-Drs. 18/12356, 23; *Roßnagel u.a.* 2018, 12.
185 S. z. B. *Roßnagel u.a.* 2018, 12; s. zur faktischen Unmöglichkeit, dies allein den staatlichen Gerichten zu überlassen *Lang*, AöR 2018, 220 (239 ff.).
186 § 11 Abs. 1, Abs. 2 RStV für den Programmauftrag der öffentlich-rechtlichen Rundfunkanstalten.
187 S. hierzu auch *Dörr/Natt*, ZUM 2014, 829; *Paal/Hennemann*, ZRP 2017, 76 (77 ff.); *Pille* 2016, 346 ff.
188 Auf beachtenswerte Stolperfallen weisen *Schulz/Dreyer* 2018, 16 ff. hin.
189 Näher dazu *Paal*, MMR 2018, 567 (570).

Insgesamt enthält der Entwurf viele brauchbare Ansätze. Die vorgeschlagenen neuen Vorschriften für Medienintermediäre betreffen u.a. funktionsbezogene Transparenzverpflichtungen, beispielsweise in § 53d MStV-E hinsichtlich der zentralen Kriterien für die Aggregation, Selektion und Präsentation von Inhalten. Denkbar wäre es auch, der bisher mangelnden Transparenz darüber, inwieweit und nach welchen Kriterien Inhalte automatisiert detektiert und automatisiert über Maßnahmen wie Sperrung und Löschung von Inhalten sowie Nutzerkonten entschieden wird, mit Transparenzpflichten abzuhelfen. Hinsichtlich solcher Transparenzpflichten müssen die Grenzen zu schützenswerten Geheimhaltungsinteressen der Anbieter eingehalten werden.

Grundsätzlich sinnvoll ist auch eine Kennzeichnungspflicht für Social Bots, wie sie die §§ 55 Abs. 3, 53d Abs. 4 MStV-E enthalten.[190] Die Kennzeichnung von Social Bots ist im Interesse des freien, demokratischen Meinungsbildungsprozesses und des Persönlichkeitsschutzes begrüßenswert. Eine Kennzeichnungsverpflichtung und die legale Verwendung von Social Bots könnte auch das Bewusstsein der Nutzenden für die Frage steigern, ob eine Nachricht von einem Menschen oder einem Algorithmus stammt.[191] Die Herstellung von Transparenz im Hinblick auf die Herkunft und Erzeugungsweise von Inhalten ist auch zum Schutz der Informationsfreiheit und der integren gesellschaftlichen Kommunikation erforderlich und angemessen.[192] Der konkrete Vorschlag liefert – die technische Machbarkeit der Kennzeichnung vorausgesetzt[193] – ein prinzipiell geeignetes Grundgerüst zur Verpflichtung von Bot-Verwendern und Netzwerkbetreibern als Regelungsadressaten, das nur kleinere Nachbesserungen erfordert. Insbesondere im Hinblick auf Beschwerdeverfahren bei Falschkennzeichnungen fehlen bislang Vorgaben, obgleich sich ein Anspruch auf Rückgängigmachung bereits aus den Nutzungsverträgen ergeben dürfte. Die Durchsetzungsmöglichkeiten von Aufsichtsmaßnahmen sollten im Vorfeld eingehend eruiert werden, um eine symbolische Gesetzgebung zu vermeiden.

Im Hinblick auf den Umgang mit Desinformationen und Social Bots erscheint ein Regulierungsmix sinnvoll. Wirksame Selbstregulierungsmaßnahmen der Social Networks bleiben ein wichtiger Bestandteil, nicht zuletzt, da sie sowohl in ihrer territorialen Reichweite als auch in ihrer Intensität über gesetzliche Verpflichtungen hinausgehen können. Beispielsweise wären etwaige gesetzliche Pflichten zur Überprüfung von Inhalten auf ihren

190 S. hierzu ausführlich *Löber/Roßnagel*, MMR 2019, 493 (497f.).
191 *Hegelich* 2016, 7; *Gräfe*, PinG 2019, 5 (11).
192 *Löber/Roßnagel*, MMR 2019, 493 (494 ff.).
193 S. Kap. 4.B.V.

Wahrheitsgehalt kaum in Einklang zu bringen mit dem unionsrechtlichen allgemeinen Überwachungsverbot.[194] So können die Social Networks z. B. in klar formulierten Nutzungsbedingungen die Verwendung von Social Bots untersagen und für den Fall des Zuwiderhandelns schärfere Maßnahmen als die Bot-Kennzeichnung, insbesondere die Löschung des betroffenen Accounts, durchsetzen. Im Falle einer gesetzlichen Regelung für Social Networks ist es naheliegend, dass sie, wie auch im Zuge des NetzDG zu beobachten ist, die Durchsetzung ihrer eigenen Regeln verbessern wird.[195] Für wirksame Selbstregulierungsmaßnahmen sollte beachtet werden, dass die Selbstregulierung der Netzwerkbetreiber nur dann zum Erfolg für das Allgemeinwohl und die Grundrechte der jeweils Betroffenen führt, wenn diese Selbstregulierungsmaßnahmen in einem gesetzlichen Rahmen stattfinden, der Gemeinwohlorientierung und Fairness sicherstellt.[196]

III. Regeln für Anbieter von Telemedien

Zu erwägen sind Sanktionsmöglichkeiten und -mechanismen für besonders schwere Verstöße gegen die Sorgfaltspflichten für journalistisch-redaktionelle Telemedien. Hierfür bietet sich eine Erweiterung der Befugnisse der Landesmedienanstalten an, um Richtigstellungen, Entfernungen oder andere Maßnahmen umsetzen zu können.[197] § 59 Abs. 3 RStV könnte dementsprechend auch für Verstöße gegen § 54 Abs. 2 RStV geöffnet werden.

Um die Weiterverbreitung von Falschinformationen einzudämmen und um den Persönlichkeitsschutz zu stärken, sollten Sorgfalts- und Recherchepflichten auch für „journalistische Laien" erwogen werden, die sich auf allgemein zugänglichen Webportalen äußern, sodass potenziell Unterlassungs- und Schadensersatzansprüche ausgelöst werden können.[198] Die Ausweitung von Sorgfaltspflichten sollte jedoch auf rechtswidrige Inhalte begrenzt sein und richterrechtlich fortgebildet werden.

Weiterhin ist zu prüfen, wie das Aufgreifen von manipulierten Stimmungsbildern gerade in „etablierten" Medienöffentlichkeiten verhindert werden kann.[199] So könnten Sorgfaltsmaßstäbe zum Umgang mit „Hashtag-

194 *Mengden* 2018, 268f.; *Peifer*, CR 2017, 809 (813).
195 S. dazu *Löber/Roßnagel*, MMR 2019, 71.
196 S. hierzu *Roßnagel* 2018, § 5 Rn. 227f.; *ders*. 2003, Kap. 3.4 Rn. 52 ff.
197 Vgl. *Bader u.a.* 2018; s. auch *Die Medienanstalten* 2018, 12.
198 *Peifer,* AfP 2015, 193 (195); *ders.*, JZ 2013, 853 (869).
199 *Dankert/Dreyer*, K&R 2017, 73 (78).

Öffentlichkeiten“ und für Berichte über Stimmungsbilder in Social Networks in Betracht gezogen werden, die explizit Social Bots und die Gefahr von durch diese verzerrten Stimmungsbilder aufgreifen. Diese könnten im Pressekodex verankert werden.[200]

E. Fazit

Im Umgang mit Desinformationen bieten sich verschiedene, zusammenwirkende Ansätze aus dem Bereich der staatlichen, der Ko- und Selbstregulierung an. Auch sind technische und regulatorische Maßnahmen zu kombinieren. Hoheitliche Regulierung sollte vor allem zum Schutz beeinträchtigter Drittrechte forciert werden. Die Meinungsfreiheit des Grundgesetzes setzt bei nicht rechtswidrigen Äußerungen auf die Kraft der gesellschaftlichen Auseinandersetzung. Die Aufgaben des Staats sind hier sehr begrenzt und richten sich auf die Aufrechterhaltung der Bedingungen für die Selbstorganisation gesellschaftlicher Kommunikation.[201] Die Regulierungsansätze sollten zudem daran gemessen werden, ob sie die Meinungsvielfalt und damit auch die freie Meinungsbildung begünstigen.[202] Die weitergehende Regulierung markt- und wirkmächtiger Online-Plattformen ist geboten. Die zunehmende Monopolisierung und Zentralisierung sowie damit korrelierende Informations- und Wissensasymmetrien unterstützen die Verbreitung von Desinformationen. Sie sind daher durch Verpflichtung der Social Networks zu Transparenz und Verantwortungswahrnehmung sowie durch hoheitliche Rahmensetzung und Eingriffsmaßnahmen auszugleichen. Wichtig sind aber insbesondere auch Initiativen anderer Akteure, die in der Zusammenfassung dieses Buchs aufgezeigt werden. Denn Desinformation ist ein vielschichtiges, vor allem gesellschaftliches Problem. Der wirksame Umgang mit diesem Problem – eine Problemlösung anzustreben wäre wohl vermessen – erfordert die Einbeziehung einer Vielzahl von Akteuren auf verschiedenen gesellschaftlichen Ebenen. Das Recht kann hierzu nur einen begrenzten Beitrag liefern.

200 S. *Dankert*, in: Hoffmann-Riem 2018, 164.
201 *Ladeur*, in: Eifert/Gostomzyk 2018, 169 (176).
202 BVerfGE 57, 295 (323) – Drittes Rundfunkurteil = NJW 1981, 1774 (1776); *Paal/Hennemann*, ZRP 2017, 76 (77).

Literaturverzeichnis zu Kapitel 5

Abel, Ralf B., Automatisierte Entscheidungen im Einzelfall gem. Art. 22 DS-GVO – Anwendungsbereich und Grenzen im nicht-öffentlichen Bereich, ZD 2018, 304.

Bader, K., Jansen, C., Johannes, P. C., Krämer, N., Kreutzer, M., Löber, L. I., Rinsdorf, L., Rösner, L. & Roßnagel, A. (2018). Desinformationen aufdecken und bekämpfen: Handlungsempfehlungen. Policy-Paper. Hrsg.: Fraunhofer-Institut für System- und Innovationsforschung ISI. Verfügbar unter https://www.forum-privatheit.de/wp-content/uploads/Policy-Paper-DORIAN-Desinformation-aufdecken-und-bekaempfen-1.pdf

Ballhausen, Miriam/Roggenkamp, Jan Dirk, Personenbezogene Bewertungsplattformen, K&R 2008, 403.

Benda, Ernst/Maihofer, Werner/Vogel, Hans-Jochen (Hrsg.), Handbuch des Verfassungsrechts der Bundesrepublik Deutschland – Teil 1, 2. Aufl. Berlin 1995, Reprint aus 2011.

Beurskens, Michael, „Hate-Speech" zwischen Löschungsrecht und Veröffentlichungspflicht, NJW 2018, 3418.

Brings-Wiesen, Tobias, „Meinungskampf mit allen Mitteln und ohne Regeln", JuWiss vom 30.11.2016, https://www.juwiss.de/93-2016/.

Bull, Hans Peter, Über die rechtliche Einbindung der Technik, Der Staat 58 (2019), 57.

Dankert, Kevin, Verfälschung von Datenbeständen durch Social Bots, in: Hoffmann-Riem, Wolfgang (Hrsg.), Big Data - Regulative Herausforderungen, Baden-Baden 2018, 157.

Dankert, Kevin/Dreyer, Stephan, Social Bots – Grenzenloser Einfluss auf den Meinungsbildungsprozess?, K&R 2017, 73.

Deutscher Bundestag Parlamentsnachrichten (Hrsg.), NetzDG auf dem Prüfstand Ausschuss Digitale Agenda/Ausschuss – 18.10.2018 (hib 781/2018), Berlin 2018.

Die Medienanstalten (Hrsg.), Stellungnahme der Medienanstalten zum Diskussionsentwurf der Rundfunkkommission der Länder zu den Bereichen Rundfunkbegriff, Plattformregulierung und Intermediäre „Medienstaatsvertrag", Berlin 2018.

Dörr, Dieter/Natt, Alexander, Suchmaschinen und Meinungsvielfalt – Ein Beitrag zum Einfluss von Suchmaschinen auf die demokratische Willensbildung, ZUM 2014, 829.

Dreier, Horst (Hrsg.), Grundgesetz-Kommentar Band I, 3. Aufl. Tübingen 2013.

Drexl, Josef, Bedrohung der Meinungsvielfalt durch Algorithmen, ZUM 2017, 529.

Dreyer, Stephan/Schulz, Wolfgang, Was bringt die Datenschutz-Grundverordnung für automatisierte Entscheidungssysteme? Potenziale und Grenzen der Absicherung individueller, gruppenbezogener und gesellschaftlicher Interessen, Gütersloh 2018.

Eifert, Martin, Rechenschaftspflichten für soziale Netzwerke und Suchmaschinen – Zur Veränderung des Umgangs von Recht und Politik mit dem Internet, NJW 2017, 1450.

Elsaß, Lennart/Labusga, Jan-Hendrik/Tichy, Rolf, Löschungen und Sperrungen von Beiträgen und Nutzerprofilen durch die Betreiber sozialer Netzwerke, CR 2017, 234.

Enders, Peter, Einsatz künstlicher Intelligenz bei juristischer Entscheidungsfindung, JA 2018, 721.

Epping, Volker/Hillgruber, Christian (Hrsg.), BeckOK Grundgesetz, 40 Aufl. München 2019.

Europäische Kommission, Aktionsplan gegen Desinformation, JOIN(2018) 36 final, Brüssel 2018.

Facebook Inc. (Hrsg.), NetzDG-Transparenzbericht – Juli 2019, Menlo Park 2019.

Faßbender, Kurt, Was darf die Satire? Bemerkungen aus der Perspektive des deutschen Verfassungsrechts, NJW 2019, 705.

Freitas, Carlos et al., An empirical study of socialbot infiltration strategies in the Twitter social network, Social Network Analysis and Mining 2016, 1.

Frisch, Wolfgang, Voraussetzungen und Grenzen staatlichen Strafens, NStZ 2016, 16.

Gersdorf, Hubertus, Hate Speech in sozialen Netzwerken, MMR 2017, 439.

Gersdorf, Hubertus/Paal, Boris P. (Hrsg.), BeckOK Informations- und Medienrecht, 24. Aufl. München 2019.

Gola, Peter (Hrsg.), DS-GVO Kommentar, 2. Aufl. München 2018.

Gounalakis, Georgios, Rechtliche Grenzen der Autocomplete-Funktion von Google, NJW 2013, 2321.

Graber, Robin/Lindemann, Thomas, Neue Propaganda im Internet. Social Bots und das Prinzip sozialer Bewährtheit als Instrumente der Propaganda, in: Sachs-Hombach, Klaus/Zywietz, Bernd (Hrsg.), Fake News, Hashtags & Social Bots, Neue Methoden populistischer Propaganda, Wiesbaden 2018, 59.

Gräfe, Hans-Christian, Webtracking und Microtargeting als Gefahr für Demokratie und Medien, PinG 2019, 5.

Grimm, Dieter, Die Meinungsfreiheit in der Rechtsprechung des Bundesverfassungsgerichts, NJW 1995, 1697.

Guggenberger, Nikolas, Das Netzwerkdurchsetzungsgesetz in der Anwendung, NJW 2017, 2577.

Hartl, Korbinian, Suchmaschinen, Algorithmen und Meinungsmacht, Wiesbaden 2017.

Häberle, Peter, Wahrheitsprobleme im Verfassungsstaat, Baden-Baden 1995.

Hegelich, Simon, Invasion der Meinungs-Roboter, in: Konrad-Adenauer-Stiftung e.V (Hrsg.), Analysen und Argumente, Berlin 2016, 1.

Heilmann, Stefan, Anonymität für User-Generated Content?, Baden-Baden 2013.

Hellgardt, Alexander, Wer hat Angst vor der unmittelbaren Drittwirkung?, JZ 2018, 901.

Hilgendorf, Eric, Beleidigung. Grundlagen, interdisziplinäre Bezüge und neue Herausforderungen, EWE 2008, 403.

Hilgendorf, Eric, Ehrenkränkungen („flaming“) im Web 2.0. Ein Problemaufriss de lege lata und de lege ferenda, ZIS 2010, 208.

Hillgruber, Christian, Die Meinungsfreiheit als Grundrecht der Demokratie. Der Schutz des demokratischen Resonanzbodens in der Rechtsprechung des BVerfG, JZ 2016, 495.

Hoffmann-Riem, Wolfgang, Freiheitsschutz in den globalen Kommunikationsinfrastrukturen, JZ 2014, 53.

Hoffmann-Riem, Wolfgang, Verhaltenssteuerung durch Algorithmen – Eine Herausforderung für das Recht, AöR 2017, 1.

Hölig, Sascha/Hasebrink, Uwe, Reuters Institute Digital News Report 2019 – Ergebnisse für Deutschland, Hamburg 2019.

Holznagel, Bernd, Meinungsbildung im Internet, NordÖR 2011, 205.

Holznagel, Bernd, Phänomen „Fake News“ – Was ist zu tun?, MMR 2018, 18.

Holznagel, Daniel, Notice and Take-Down-Verfahren als Teil der Providerhaftung, Tübingen 2013.

Holznagel, Daniel, Overblocking durch User Generated Content (UGC) – Plattformen: Ansprüche der Nutzer auf Wiederherstellung oder Schadensersatz?, CR 2018, 369.

Hoven, Elisa/Krause, Melena, Die Strafbarkeit der Verbreitung von „Fake News“, JuS 2017, 1167.

Ingold, Albert, Desinformationsrecht: Verfassungsrechtliche Vorgaben für staatliche Desinformationstätigkeit, Berlin 2011.

Ingold, Albert, Propaganda als Herausforderung des Kommunikationsrechts, in: Oppelland, Torsten (Hrsg.), Propaganda als (neue) außen- und sicherheitspolitische Herausforderung, Berlin 2018, 81.

Isensee, Josef/Kirchhof, Paul (Hrsg.), Handbuch Staatsrecht, Band VII, 3. Aufl. Heidelberg 2009.

Jarass, Hans D./Pieroth, Bodo (Hrsg.), Grundgesetz für die Bundesrepublik Deutschland – Kommentar, 15. Aufl. München 2018.

Kersten, Jens, Anonymität in der liberalen Demokratie, JuS 2017, 193.

Kind, Sonja u.a., Social Bots TA-Vorstudie, Berlin 2017.

Koreng, Ansgar, Hate-Speech im Internet – Eine rechtliche Annäherung, KriPoZ 2017, 151.

Krischker, Sven, „Gefällt mir“, „Geteilt“, „Beleidigt“? – Die Internetbeleidigung in sozialen Netzwerken, JA 2013, 488.

Kušen, Ema/Strembeck, Mark, Something draws near, I can feel it: An analysis of human and bot emotion-exchange motifs on Twitter, Online Social Networks and Media, Vol. 10-11 (2019), 1.

Lackner/Kühl, StGB Kommentar, hrsg. von Kühl, Kristian/Heger, Martin, 29 Aufl. München 2018.

Ladeur, Karl-Heinz, Netzwerkrecht als neues Ordnungsmodell des Rechts, in: Eifert, Martin/Gostomzyk, Tobias (Hrsg.), Netzwerkrecht. Die Zukunft des NetzDG und seine Folgen für die Netzwerkkommunikation, Baden-Baden 2018, 169.

Lang, Andrej, Netzwerkdurchsetzungsgesetz und Meinungsfreiheit. Zur Regulierung privater Internet-Intermediäre bei der Bekämpfung von Hassrede, AöR 2018, 220.

Lausen, Matthias, Unmittelbare Verantwortlichkeit des Plattformbetreibers, ZUM 2017, 278.

Lazer, David M. J. et al., The science of fake news, Science 2018, Vol. 359, Issue 6380, 1094.

Libertus, Michael, Rechtliche Aspekte des Einsatzes von Social Bots de lege lata und de lege ferenda, ZUM 2018, 20.

Liesching, Marc, Die Durchsetzung von Verfassungs- und Europarecht gegen das NetzDG – Überblick über die wesentlichen Kritikpunkte, MMR 2018, 26.

Löber, Lena Isabell/Roßnagel, Alexander, Das Netzwerkdurchsetzungsgesetz in der Umsetzung, MMR 2019, 71.

Löber, Lena Isabell/Roßnagel, Alexander, Kennzeichnung von Social Bots – Transparenzpflichten zum Schutz integrer Kommunikation, MMR 2019, 493.

Lüdemann, Jörn, Grundrechtliche Vorgaben für die Löschung von Beiträgen in sozialen Netzwerken, MMR 2019, 279.

Mafi-Gudarzi, Nima, Desinformation: Herausforderung für die wehrhafte Demokratie, ZRP 2019, 65.

Martini, Mario/Nink, David, Wenn Maschinen entscheiden… Persönlichkeitsschutz in vollautomatisierten Verwaltungsverfahren, NVwZ 2017, 681.

Maunz/Dürig, Grundgesetz-Kommentar, hrsg. von Herzog, Roman/Scholz, Rupert/Herdegen, Matthias/Klein, Hans H., 86 EL., München 2019.

Mengden, Martin, Zugangsfreiheit und Aufmerksamkeitsregulierung. Zur Reichweite des Gebots der Gewährleistung freier Meinungsbildung am Beispiel algorithmengestützter Zugangsdienste im Internet, Tübingen 2018.

Michl, Michael, Situativ staatsgleiche Grundrechtsbindung privater Akteure, JZ 2018, 910.

Milker, Jens, »Social-Bots« im Meinungskampf, ZUM 2017, 216.

Müller-Franken, Sebastian, Netzwerkdurchsetzungsgesetz: Selbstbehauptung des Rechts oder erster Schritt in die selbstregulierte Vorzensur? – Verfassungsrechtliche Fragen, AfP 2018, 1.

Noelle-Neumann, Elisabeth, Öffentliche Meinung: Die Entdeckung der Schweigespirale, Berlin 1996.

Paal, Boris P., Vielfaltssicherung bei Intermediären – Fragen der Regulierung von sozialen Netzwerken, Suchmaschinen, Instant-Messengern und Videoportalen, MMR 2018, 567.

Paal, Boris P./Hennemann, Moritz, Meinungsbildung im digitalen Zeitalter Regulierungsinstrumente für einen gefährdungsadäquaten Rechtsrahmen, JZ 2017, 641.

Paal, Boris P./Hennemann, Moritz, Meinungsvielfalt im Internet – Regulierungsoptionen in Ansehung von Algorithmen, Fake News und Social Bots, ZRP 2017, 76.

Papier, Hans-Jürgen, Rechtsstaatlichkeit und Grundrechtsschutz in der digitalen Gesellschaft, NJW 2017, 3025.

Peifer, Karl-Nikolaus, Die zivilrechtliche Verteidigung gegen Äußerungen im Internet, AfP 2015, 193.

Peifer, Karl-Nikolaus, Fake News und Providerhaftung. Warum das NetzDG zur Abwehr von Fake News die falschen Instrumente liefert, CR 2017, 809.

Peifer, Karl-Nikolaus, Persönlichkeitsrechte im 21. Jahrhundert – Systematik und Herausforderungen, JZ 2013, 853.

Petruzzelli, Michelle, Bewertungsplattformen, Überdehnung der Meinungsfreiheit zu Lasten der Betroffenen vs. gerechtfertigte Einschränkung zu Lasten der Bewertenden, MMR 2017, 800.

Peukert, Alexander, Gewährleistung der Meinungs- und Informationsfreiheit in sozialen Netzwerken, MMR 2018, 572.

Pille, Jens-Ullrich, Meinungsmacht sozialer Netzwerke, Baden-Baden 2016.

Raue, Benjamin, Meinungsfreiheit in sozialen Netzwerken – Ansprüche von Nutzern sozialer Netzwerke gegen die Löschung ihrer Beiträge, JZ 2018, 961.

Ross, Björn et al., Are social bots a real threat?, EJIS 2019, 1.

Roßnagel, Alexander, Konzepte der Selbstregulierung, in: ders. (Hrsg.), Handbuch Datenschutzrecht, München 2003.

Roßnagel, Alexander, Der künftige Anwendungsbereich der Fernsehrichtlinie, in: Institut für Europäisches Medienrecht (EMR) (Hrsg.), Die Zukunft der Fernsehrichtlinie, Baden-Baden 2005, 35.

Roßnagel, Alexander, Beck'scher Kommentar zum Recht der Telemediendienste, Kommentar zum TMG, SigG, SigV, JMStV, BGB, VwVfG, ZPO, München 2013.

Roßnagel, Alexander, Verhaltensregeln, in: ders. (Hrsg.), Das neue Datenschutzrecht, Baden-Baden 2018, § 5 Rn. 193.

Roßnagel, Alexander u.a., Policy Paper – Das Netzwerkdurchsetzungsgesetz, Karlsruhe 2018.

Roßnagel, Alexander, Datenschutz in der Forschung. Die neuen Datenschutzregelungen in der Forschungspraxis von Hochschulen, ZD 2019, 157.

Rühl, Ulli, Tatsachenbehauptungen und Wertungen, AfP 2000, 17.

Sachs, Michael (Hrsg.), Grundgesetz-Kommentar, 8. Aufl. München 2018.

Sängerlaub, Alexander, Feuerwehr ohne Wasser? Möglichkeiten und Grenzen des Fact-Checkings als Mittel gegen Desinformation, Berlin 2018.

Schiff, Alexander, Meinungsfreiheit in mediatisierten digitalen Räumen, MMR 2018, 366.

Schliesky, Utz u.a., Schutzpflichten und Drittwirkung im Internet, Baden-Baden 2014.

Schmitz-Remberg, Florian J., Verständigung und positive Generalprävention, Düsseldorf 2014.

Schönke, Adolf/Schröder, Horst (Begr.), Strafgesetzbuch Kommentar, 30. Aufl. München 2019.

Schröder, Meinhard, Rahmenbedingungen der staatlichen Regulierung von Social Bots, DVBl 2018, 465.

Schulz, Wolfgang/Dreyer, Stephan, Stellungnahme zum Diskussionsentwurf eines Medienstaatsvertrags der Länder, Hamburg 2018.

Seckelmann, Margrit, Evaluation und Recht, Tübingen 2018.

Simitis, Spiros/Hornung, Gerrit/Spiecker gen. Döhmann, Indra (Hrsg.), Datenschutzrecht – DSGVO mit BDSG, 1. Aufl. Baden-Baden 2019.

Spickhoff, Andreas (Hrsg.), Medizinrecht, 3. Aufl. München 2018.

Spindler, Gerald, Löschung und Sperrung von Inhalten aufgrund von Teilnahmebedingungen sozialer Netzwerke, CR 2019, 238.

Stark, Birgit/Magin, Melanie/Jürgens, Pascal, Politische Meinungsbildung im Netz: Die Rolle der Informationsintermediäre, UFITA 2018, 103.

Steinbach, Armin, Meinungsfreiheit im postfaktischen Umfeld, JZ 2017, 653.

Steinbach, Armin, Social Bots im Wahlkampf, ZRP 2017, 101.

Stern, Klaus/Becker, Florian (Hrsg.), Grundrechte-Kommentar, 3. Aufl. Köln 2018.

Thielbörger, Pierre, Propaganda-blindes Völkerrecht?, in: Oppelland, Torsten (Hrsg.), Propaganda als (neue) außen- und sicherheitspolitische Herausforderung, Berlin 2018, 63.

Twitter Inc., Twitter Netzwerkdurchsetzungsgesetzbericht: Januar – Juni 2019, San Francisco 2019.

Ulrich, Franz, Haftung des Bewertungsportals für zu eigen gemachte Äußerungen Dritter, AfP 2017, 316.

v. Mangoldt/Klein/Starck Grundgesetz-Kommentar, hrsg. von Voßkuhle, Andreas/Huber, Peter Michael, Band I, 7 Aufl. München 2018.

Vosoughi, Soroush/Roy, Deb/Aral, Sinan, The spread of true and false news online, Science 2018, Vol. 359, Issue 6380, 1146.

Wandtke, Arthur Axel, Persönlichkeitsschutz versus Internet, MMR 2019, 142.

Woger, Hans-Christian/Männig, Annina Barbara, „Hate Speech" – Eine rechtliche Einordnung und das Netzwerkdurchsetzungsgesetz zur Bekämpfung von Hassrede im Internet, PinG 2017, 233.

Wolff, Stefan/Brink, Heinrich Amadeus (Hrsg.), BeckOK Datenschutzrecht, 27. Aufl. München 2019.

YouTube LLC, Entfernungen von Inhalten aus YouTube auf der Grundlage des NetzDG - 1. Januar 2019 - 30. Juni 2019, San Bruno 2019.

Kapitel 6: Handlungsempfehlungen

Autoren:

Dr. Carolin Jansen
Paul Christopher Johannes
Prof. Dr. Nicole Krämer
Dr. Michael Kreutzer
Lena Isabell Löber
Prof. Dr. Lars Rinsdorf
Prof. Dr. Alexander Roßnagel
Dr. Leonie Schaewitz

Wie im Fazit des vorangegangenen Kapitels dargelegt, handelt es sich bei digitaler Desinformation um ein Problem, das unter Einbeziehung einer Vielzahl von Akteuren auf verschiedenen gesellschaftlichen Ebenen angegangen werden muss. Die Handlungsempfehlungen dieses Kapitels sind dementsprechend nach den Adressatengruppen strukturiert.[1]

A. Empfehlungen für Bürgerinnen und Bürger

I. Merkmale von Desinformation (wiederer)kennen und Plausibilitätschecks vornehmen

Bürgerinnen und Bürger, die sich in politische Diskurse einbringen möchten, sind auf richtige und vielfältige Informationen angewiesen. Insbesondere technikaffinen Bürgerinnen und Bürgern stehen bereits jetzt eine Reihe hilfreicher Instrumente zur Verfügung, mit deren Hilfe sie (Des-)Informationen überprüfen können.

1 Die nachfolgenden Empfehlungen korrespondieren zu den im Policy Paper „Desinformation aufdecken und bekämpfen“ (Bader et al., 2018) gegebenen. Sie erweitern das Policy Paper in zweierlei Hinsicht. Auf Basis der zwischenzeitlich erlangten Forschungsergebnisse werden neue Empfehlungen ergänzt. Zudem enthält die nachfolgende Darstellung die wissenschaftlichen Belege, auf denen sie fußen.

1. Webseiten

Bezogen auf Webseiten lässt sich prüfen, ob die Website über ein Impressum verfügt und dies ausreichende Informationen über die presserechtlich verantwortlichen Betreiber enthält. Falls die Webseite in verschiedenen Fact-Checking-Initiativen mehrfach erwähnt wird, dann handelt es sich um erhebliche Verdachtsmomente. In Internet-Archiven lässt sich zudem recherchieren, welche Inhalte auf jenen Webseiten in der Vergangenheit publiziert wurden (s. Kapitel 4, A.I.).

2. Bilder und Videos

Auch für Bilder und Videos gibt es erste Plaubilitätschecks, die auch Laien durchführen können: Eine Rückwärtsbildersuche liefert etwa sehr schnell Hinweise darauf, in welchem Umfeld identische oder sehr ähnliche Bilder bereits veröffentlicht wurden (s. Kapitel 4, B.III.1.). Auch Metadaten, die über Exif-Viewer erfasst werden können, lassen Rückschlüsse auf die Authentizität des Bildmaterials zu.

3. Posts

Posts auf Online-Plattformen wiederum können mit Hilfe von Graph-Suchmaschinen daraufhin untersucht werden, wie die Autorinnen und Autoren von Posts, die den Verdacht erregen, Desinformation zu enthalten, miteinander vernetzt sind. Zudem liefern häufig auch Recherchen auf Online-Plattformen Hinweise auf den politischen Hintergrund von Personen, die Desinformationen teilen (s. Kapitel 4, B.I.).

4. Darstellungsmuster

Gerade technisch weniger versierten Bürgerinnen und Bürgern hilft bei der Entlarvung von Desinformationen ein konzentrierter Blick darauf, wie manipulativ intendierte (Falsch-)Meldungen geschrieben worden sind. Diese Meldungen zeigen in der Textgestaltung typische Muster in Bezug auf die (Nicht-)Einhaltung professioneller journalistischer Standards und die Instru-

mente, die eingesetzt werden, um die Aufmerksamkeit der Nutzerinnen und Nutzer zu maximieren (s. Kapitel 2, F.II.):

Vorsicht ist etwa geboten, wenn Meldungen besonders reißerisch dargestellt werden: Im Vergleich zu Meldungen, die von professionellen journalistischen Nachrichtenmedien verbreitet werden, weisen Desinformationen auf deutschsprachigen Portalen im Durchschnitt deutlich höhere Anteile nutzermaximierender Merkmale auf – wie etwa Sensationalismus oder die Skandalisierung von Ereignissen.

Der Grad der Nutzermaximierung hängt auch von dem behandelten Thema ab (s. Kapitel 2, F.III.). Hierzu beispielhaft ein wiederholt angetroffenes Muster: Eine umfassende Maximierung war zum Zeitpunkt der Erhebung von DORIAN häufiger anzutreffen, wenn gleichzeitig über Migration *und* innere Sicherheit berichtet wird, als wenn nur isoliert über eines der beiden Themen berichtet wird.

Weitere Beispiele (s. Kapitel 2, F.II.) sind unpräzise Überschriften, ein fehlender nachrichtlicher Einstieg in das Thema oder eine mangelhafte Gliederung entlang der Relevanz einzelner Informationen. Schließlich ist die Prüfung der Plattform, von der eine Meldung stammt, wichtig.

II. Die eigene Filterblase verlassen

1. Unterschiedliche Quellen nutzen

Eine zentrale Strategie für Bürgerinnen und Bürger, die sich nicht von Desinformation täuschen lassen möchten, ist ein vielfältiges Medienrepertoire, das sich aus unterschiedlichen Quellen zusammensetzt. Insbesondere durch das Lesen von verschiedenartigen Posts aus unterschiedlichen Quellen auf Online-Plattformen können Nutzende Informationen außerhalb ihrer Filterblase erhalten, ohne dass ihre Weltanschauung dabei notwendigerweise konkret angegriffen wird. Zusätzlich führt eine vielfältige Nutzung von Onlinemedien dazu, dass die Fehlwahrnehmung eines Sachverhalts durch Personen, die Desinformation gelesen haben, reduziert werden kann. Darüber hinaus könnte das Bewusstmachen des Bestätigungsfehlers helfen, Desinformation – auch wenn sie die eigene Einstellung bestätigen würde – eher zu erkennen.

Diese Empfehlung setzt bei den typischen Vorgehensweisen an, die Individuen bei der Nutzung von Nachrichten anwenden: Sie verarbeiten und bewerten neue Informationen basierend auf ihren vorherigen Einstellungen.

Dabei wenden sie sich in ihrer Auswahl eher den Nachrichten zu, die ihren Ansichten und politischen Einstellungen entsprechen. Da sie motiviert sind, ihre Ansichten zu verteidigen, neigen sie zudem dazu, Informationen zu akzeptieren, die mit ihrer Einstellung übereinstimmen und davon abweichende zu vermeiden oder anzuzweifeln.

Es ist daher anzunehmen, dass Desinformation eher von Personen mit einer deckungsgleichen Einstellung zum Thema der Nachricht geglaubt wird. Empirische Untersuchungen konnten einen signifikanten positiven Zusammenhang zwischen der wahrgenommenen Glaubwürdigkeit einer fabrizierten Falschmeldung und der Einstellung zum Thema der Nachricht feststellten (s. Kapitel 3, B.I.1.). Eine zusätzliche Gefahr geht von einseitigem Nachrichtenkonsum zu einem Thema aus (wenn man durch den Bestätigungsfehler bestimmte Nachrichtenkanäle/Sender bevorzugt). Das wiederholte Lesen und Wahrnehmen von einer Falschinformation führt dazu, dass diese für glaubwürdiger gehalten wird.

III. Desinformation melden und Freunde ansprechen

1. Verbreitung eindämmen

Um Falschinformationen zu bekämpfen, scheint es wirkungsvoller zu sein, die Verbreitung von vorneherein einzudämmen, anstatt im Nachhinein zu versuchen, diese zu korrigieren (s. Kapitel 3, B.II.1.). Hinsichtlich der Verbreitung hat sich gezeigt, dass vor allem Menschen und nicht Social Bots zur Weiterleitung von Desinformation beitragen (Vosoughi et al., 2018). Dies wird insbesondere darauf zurückgeführt, dass Menschen oft Nachrichten weiterleiten, ohne sich intensiv mit deren Inhalt auseinanderzusetzen. Die Entscheidung, eine Nachricht zu teilen, fällt vielmehr in der Regel auf Basis des Informationsgehalts und der sprachlichen Gestaltung von Überschrift und Teasern (Gabielkov et al., 2016; Vosoughi et al., 2018).

2. Desinformation nicht weiterleiten, bei Verdacht auf strafbare Inhalte der jeweiligen Distributionsplattform melden

Ein wichtiger Weg, Desinformation einzudämmen, besteht somit darin, dafür zu sorgen, dass Desinformationen nicht weitergeleitet werden. Personen, die Desinformation auf Online-Plattformen entdecken oder den Verdacht haben,

dass es sich um Desinformation handelt, sollten die Meldung daher nicht an ihr Umfeld weiterleiten, sondern der jeweiligen Distributionsplattform melden, sofern eine solche Funktionalität – wie etwa bei Facebook – angeboten wird.

Mit der Meldefunktion können sie auch (strafbare) Inhalte melden, die sie nicht selbst betreffen. Die Online-Plattformen müssen nach den Regelungen der begrenzten Hostproviderhaftung nicht proaktiv nach Gesetzesverstößen suchen. Erst wenn sie Kenntnis von den rechtswidrigen Inhalten erlangen, zumeist durch Meldungen von Nutzenden, werden die Prüf- und Löschpflichten aktiviert (s. Kapitel 5, C.II.).

3. Keine Desinformation wiederholen

Eine Weiterleitung mit gleichzeitigem Warnhinweis empfiehlt sich nicht, da noch unklar ist, inwieweit die Falschnachricht trotz vorhandener Kennzeichnung nicht als falsch abgespeichert wird (s. Kapitel 3, B.II.2.). Ferner können Weiterleitung und Kommentierung dazu beitragen, dass die Nachricht über Auswahl- und Verbreitungs-Algorithmen erfolgreicher wird. Bürgerinnen und Bürger, die die Qualität politischer und gesellschaftlicher Diskurse auf Online-Plattformen steigern wollen, indem sie Falschmeldungen sorgfältig recherchierte Fakten entgegenstellen, sollten darauf achten, dass in ihren Posts die ursprüngliche, falsche Nachricht nicht als Link enthalten ist.

4. Hinweis auf Desinformation

Außerdem kann es hilfreich sein, Freunde und Bekannte, die Desinformation geteilt haben, direkt darauf hinzuweisen. Ein Vertrauensverhältnis zwischen den Nutzenden kann bei der Akzeptanz von Korrekturdarstellungen von Vorteil sein: Korrekturen auf Twitter sind bedeutend häufiger erfolgreich, wenn die beteiligten Personen miteinander vernetzt waren (Margolin et al. 2018). Demnach ist es empfehlenswert, innerhalb seines persönlichen digitalen Netzwerks proaktiv darauf hinzuweisen, wenn eine vertraute Person einer Falschinformation aufsitzt und sie verbreitet. Damit dies gelingen kann, sollte man darauf achten, auch die Informationen besonders zu prüfen, die von Freunden kommen, denn prinzipiell kann man davon ausgehen, dass auf Basis des Vertrauens, das man in Freunde hat, im persönlichen Umfeld eher seltener Fact-Checking betrieben wird (s. Kapitel 3, D.).

B. Empfehlungen an Medienunternehmen, die sich zum Pressekodex bekennen

I. Sorgfalt vor Schnelligkeit, Sachlichkeit vor Leseanreiz

1. Professionelle Standards wahren

Um bei den Nutzenden glaubwürdiger zu wirken, übernehmen die Autorinnen und Autoren von Desinformation häufig professionelle Regeln für das Schreiben von Nachrichten. Allerdings hat die Forschung (s. Kapitel 2, C.) gezeigt, dass diese Mimikry nur teilweise gelingt: Gerade komplexere journalistische Regeln, die die Argumentationsstruktur betreffen, werden von den meisten Verfassern von Fake News nicht eingehalten. Nachrichtenmedien sollten professionelle Standards der Nachrichtenerstellung noch konsequenter einhalten – dies wird sie noch klarer von Desinformationsangeboten abgrenzen.

2. Angaben zu Fakten prüfen

Hierzu gehört zum einen, der sorgfältigen Prüfung von Fakten im Zweifel den Vorzug zu geben vor einer möglichst schnellen Veröffentlichung einer Nachricht – auch im gerade in dieser Hinsicht harten Wettbewerbsumfeld auf Online-Märkten. Dies zahlt sich mittelfristig in einer höheren Glaubwürdigkeit aus. Jede fehlerhafte Meldung schwächt die Glaubwürdigkeit von seriösen Medien bei Nutzenden, faktenbasierte Berichterstattung stärkt sie (s. Kapitel 2, F.II.).

3. Sachliche Meldungen

Auch bei professionellen journalistischen Meldungen ist es legitim und richtig, mit sprachlichen Mitteln Aufmerksamkeit für relevante Inhalte zu generieren – gerade auch in Teasern, die auf Kommunikationsplattformen verbreitet werden. Angesichts des massiven Einsatzes dieser Instrumente auf Desinformationsplattformen empfiehlt sich im Sinne einer Abgrenzungsstrategie auch hier Zurückhaltung – im Zweifel sollten sich Redaktionen für die sachlichere Variante entscheiden (s. Kapitel 2, C.).

4. Offene Fehlerkultur

Bei aller journalistischen Sorgfalt lassen sich Fehler nicht vermeiden. Ein offener Umgang mit Fehlern und eine grundsätzlich hohe Transparenz hinsichtlich der redaktionellen Abläufe und den Qualitätssicherungsmechanismen sind daher mehr denn je geboten, um die nach wie vor hohe Glaubwürdigkeit, die etablierte Nachrichtenmedien in Deutschland genießen, auf diesem Niveau zu halten (Sängerlaub et al., 2018: 12).

Gerade eine offene Fehlerkultur findet sich auf Desinformationsportalen nicht. Und so gut es diesen Portalen gelingt, offensichtliche Fehler oder Mängel zu vermeiden, ist es für sie schwierig, professionelle journalistische Standards in einer Vielzahl von Dimensionen zu erfüllen, von geeigneten Schlagzeilen bis hin zur Konsistenz der präsentierten Fakten. Einige von ihnen versuchen, diese Defizite durch den Verweis auf glaubwürdigere Quellen auszugleichen.

II. Technische Unterstützung zur Aufdeckung von Desinformation nutzen und vorantreiben

Redakteurinnen, die den Wahrheitsgehalt einer Meldung untersuchen, können inzwischen auf eine Vielzahl von technischen Werkzeugen und Dienstleistungen zurückgreifen, die sie hierbei unterstützen (s. Kapitel 4, C.). Um mittels maschinellen Lernens in der Zukunft immer bessere Ergebnisse zu erzielen, ist das Schaffen einer Trainingsgrundlage durch Medienunternehmen von großer Bedeutung. Werden große Mengen von Texten und auch anderen Medien, die als Desinformation erkannt wurden, in Datenbanken abgelegt und entsprechend kommentiert, dann wird die Erkennung immer präziser (s. Kapitel 4, B.).

III. Entlarvung und Korrektur ansprechend gestalten

Die Entlarvung von Desinformation ist eine wichtige gesellschaftliche Aufgabe, der sich zivilgesellschaftliche Akteure ebenso annehmen wie Redaktionen etablierter journalistischer Medien. Dabei ist es eine Herausforderung für Fakt-Checking-Organisationen und journalistische Aufklärungsarbeit, ansprechende Narrative zu finden, damit sich korrekte Informationen schnell und breit verteilen. Eine ansprechende Geschichte erfüllt dabei zwei Zwecke:

Zum einen kann sie dazu beitragen, dass das Interesse der Leserinnen und Leser geweckt und die korrekte Information mit höherer Wahrscheinlichkeit gelesen und weitergeleitet wird. Zum anderen kann die Verpackung in eine Geschichte (inklusive korrekter, alternativer Erklärungen und Hintergrundinformationen) dazu beitragen, dass die richtige Information fester in den Wissensbestand aufgenommen wird.

1. Ansprechende Narrative

Ansprechende Narrative sind ein zentraler Erfolgsfaktor für Fact-Checking-Aktivitäten, denn eine Korrekturmeldung muss eine vergleichbar hohe Reichweite erzielen, wie die Falschmeldung selbst, um eine vergleichbare Wirkung erzielen zu können (s. Kapitel 3, B.II.). Schlichte Gegendarstellungen verbreiten sich im Allgemeinen in Online-Netzwerken deutlich weniger als Desinformation.

2. Regeln des Widerlegens

Bei der Widerlegung von Falschinformationen sollten grundsätzliche Regeln des Widerlegens beachtet werden, wie zum Beispiel Fakten in den Vordergrund zu stellen, komplizierte Korrekturen zu vermeiden und alternative Erklärungen anzubieten. Die Herausforderung besteht darin, einfache und kurze Korrekturen zu produzieren, die in der Wahrnehmung attraktiver sind als die entsprechende Desinformation (s. Kapitel 3, B.II.).

Darüber hinaus kann die Darstellungsform einer Widerlegung dazu beitragen, dass die relevanten Informationen von der Zielgruppe aufgenommen und erinnert werden. Sowohl Texte als auch grafische Wahrheitsskalen sind in der Lage, eine Korrektur effektiv zu vermitteln. Auch kann es hilfreich sein, die korrigierenden Informationen in Form eines Videos darzustellen, da hierdurch die Aufmerksamkeit gesteigert und gleichzeitig eine mögliche Verwirrung reduziert werden kann (Young, Jamieson, Poulsen & Goldring, 2017).

C. *Empfehlungen an Online-Plattformen und Betreiber von Systemen mit Nutzer-generierten Inhalten*

I. Overblocking und Underblocking vermeiden und freiwillige Selbstkontrolle einführen

1. Verantwortung von Online-Plattformen

Wirtschaftlich wie publizistisch einflussreiche Online-Plattformen haben sich zu relevanten Foren der öffentlichen Kommunikation entwickelt, in denen Einfluss auf den öffentlichen Meinungsaustausch und die Meinungsbildung genommen wird. Hieraus resultieren eine besondere gesellschaftliche Verantwortung sowie rechtliche Pflichten, denen sie verstärkt nachkommen müssen – auch im Umgang mit Desinformation.

Die Kommunikationsplattformen sollten ihrem selbst auferlegten Anspruch der Verteidigung der Meinungsfreiheit gerecht werden, indem sie sowohl „Over-" als auch „Underblocking" möglichst effizient vermeiden. Eine Sperrung oder Löschung von Inhalten und Nutzerkonten muss sachlich gerechtfertigt und darf nicht willkürlich sein. Bei nicht offensichtlich rechts- und regelwidrigen Inhalten ist nach den Umständen des Einzelfalls dem Ersteller von Inhalten Gelegenheit zur Stellungnahme zu geben, bevor eine Maßnahme ergriffen wird. Im Umgang mit Nutzerbeschwerden wegen (unrechtmäßig) gelöschter und gesperrter Inhalte und Nutzerkonten müssen klare Regeln und Verfahren etabliert werden, welche den Erstellern von Inhalten die Möglichkeit einräumen, einen Antrag auf (erneute) Überprüfung des Inhalts zu stellen. Die Regeln und Verfahren zur Wiederherstellung von unrechtmäßig entfernten Inhalten müssen verbessert und transparenter werden (s. Kapitel 5, C.II. und D.II.). Beim Umgang mit strafbaren Inhalten auf ihren Plattformen sollten sie die Kooperation mit den Stafverfolgungsbehörden verbessern und diese Aufgabe mit systematischen Strukturen und Prozessen angehen.

2. Meinungsfreiheit beachten

Als private Unternehmen sind Kommunikationsplattformen nicht unmittelbar an Grundrechte gebunden. Daher haben sie etwas weitere Spielräume für ihre eigenen Plattformregeln. Als wirkmächtige Anbieter öffentlicher Kommunikationsräume unterliegen sie jedoch einer intensivierten mittel-

baren Drittwirkung in Bezug auf die Kommunikationsfreiheiten und den Persönlichkeitsschutz der Nutzenden.

Wegweisend dürfte die Rechtsprechung zur Reichweite der Befugnisse der Kommunikationsplattformen bei der Formulierung und Anwendung von AGB im Zusammenhang mit Hassrede sein. Da auch die Grundrechte, insbesondere die Berufsfreiheit, der Anbieter zu berücksichtigen sind, ist zu befürworten, dass für die Anbieter die Wertentscheidungen für die Meinungsfreiheit des Grundgesetzes bei der privaten Rechtssetzung richtungsweisend sind. Dies bedeutet vor allem, dass die eigenen Regeln derart ausgestaltet sein dürfen, dass sie sich im Großen und Ganzen auf einer Linie mit den Grundsätzen zur Gewährleistung der Meinungsfreiheit bewegen. Die Unternehmen dürfen folglich im Kontext von Hassrede strengere Regeln aufstellen, sodass im Einzelfall auch Äußerungen erfasst sein können, die grundsätzlich noch von der Meinungsfreiheit erfasst sind. Hierfür spricht auch, dass durch Hassrede eine Diskussion nachhaltig negativ beeinflusst werden kann, sodass andere Nutzende eingeschüchtert werden und von einer (weiteren) Beteiligung absehen. Bei „abstrakt" politischer, öffentlichkeitsbedeutsamer Rede sind strengere Regeln hingegen nicht zulässig (s. Kapitel 5, C.III.).

Desinformation darf bei Zugrundelegung der Wertentscheidung für die Meinungsfreiheit auch unter Heranziehung etwaiger AGB in der Regel nicht entfernt werden, wenn die unwahren Tatsachenbehauptungen einen starken Meinungsbezug durch die Verbindung mit Werturteilen aufweisen, Rechte Dritter nicht verletzt sind und der Inhalt nicht strafbar ist. Ein größerer Handlungsspielraum besteht hingegen bei (offensichtlich) bewusst unwahren Tatsachenbehauptungen, die nicht in Zusammenhang mit einem Werturteil verbreitet werden. Denn bewusst unwahre Tatsachenbehauptungen fallen schon gar nicht in den Schutzbereich der Meinungsfreiheit, wenn sie unter keinem denkbaren Gesichtspunkt zum grundrechtlich geschützten Prozess der Meinungsbildung beitragen können (s. Kapitel 5, A.II.).

3. Selbstregulierung

Des Weiteren ist die Einrichtung einer Selbstregulierungsstelle „Desinformation" für die großen Online-Plattformen empfehlenswert. In regelmäßigen Berichten über den Umgang mit Desinformation und die Lösch- und Sperrpraxis können die Unternehmen zeigen, dass sie sich in der gesellschaftlichen Verantwortung sehen und sich dieser stellen. Zudem könnte so eine vertiefte, faktenbasierte Diskussion über die private Regulierung, Wirkung und Be-

deutung der Online-Plattformen aus Makroperspektive geführt und zu einem besseren Verständnis der Verfahrensweise der Online-Plattformen auch aus Sicht der Nutzenden beigetragen werden. Die im „EU-Praxiskodex Desinformation" vorgesehene Selbstverpflichtung zur jährlichen Berichterstattung ist ein wichtiger Schritt hin zu mehr Transparenz. Die im Vorfeld der Europawahl 2019 durch die Unterzeichner zu veröffentlichenden Berichte wurden seitens der EU-Kommission bereits positiv aufgenommen, die Notwendigkeit anhaltenden und verstärkten Engagements der Unterzeichner jedoch betont. Weitere Berichte sowie die erste umfassende und konkrete Würdigung der in den ersten zwölf Monaten erfolgten Maßnahmen der Unterzeichner durch die EU-Kommission Ende des Jahres 2019 bleiben abzuwarten. Die Transparenzberichte im Rahmen des Netzwerkdurchsetzungsgesetzes (NetzDG) betreffen nur bestimmte strafbare Inhalte auf den Plattformen. Die Online-Plattformen sollten sich einer freiwilligen Selbstkontrolle unterwerfen, wie sie z. B. mit Erfolg auch im Bereich des Jugendmedienschutzes etabliert ist. Hier ist es aber auch Aufgabe der Politik, auf die Einrichtung und den Ausbau entsprechender Institutionen im Bereich der Online-Plattformen hinzuwirken (s. Kapitel 5, C.V.).

II. Social Bots aufspüren und Malicious Social Bots systematisch eliminieren

Die Betreiber von Plattformen haben die beste Ausgangssituation, um Social Bots aufzuspüren (s. Kapitel 4, B.V.). Dieses Privileg sollten sie – auch im Eigeninteresse – nutzen: Bei Kenntnis können sie die Profile von Social Bots kennzeichnen (s. Kapitel 6, D.II.4.).

Die Betreiber sind bei Kenntnis der Social-Bot-Profile zudem gut vorbereitet, wenn sie wirksame Gegenmaßnahmen bei massenhafter Verbreitung von Desinformation und auch weiterer unerwünschter Inhalte (die beispielsweise gegen ihre Community-Richtlinien verstoßen) durch Malicious Bots einleiten möchten.

Darüber hinaus ist es für die Wahrung ihrer eigenen Interessen bezüglich Cybersicherheit und Privatsphärenschutz essenziell diese Profile aufzuspüren, denn die Kenntnis der Bots einer Plattform stellt ein notwendiges Element im Angreifermodell der Cybersicherheitsbetrachtung der Betreiber der Plattformen dar: Mit Hilfe dieser Bots könnten ggf. skalierbare Angriffe auf die Verfügbarkeit, Integrität und Vertraulichkeit der Betreiberinfrastruktur und der gesamten Plattform lanciert werden.

Sobald Malicious Social Bots als solche überführt wurden und ihre Aktionen nachweislich gegen Richtlinien des Plattformbetreibers verstoßen, können (und sollten) sie unmittelbar gelöscht oder gesperrt werden (s. Kapitel 5, C.I.). Für diese Aufgabe sollten die Betreiber Strukturen und systematische Prozesse vorsehen.

D. Empfehlungen für Politik und Gesetzgebung

I. Zivilgesellschaftliche Akteure (Medienbildung, Faktenchecker) unterstützen

1. Bildungsaufgaben

Die Motivation und Fähigkeit von Online-Nutzenden zum elaborierten Umgang mit (Des-)Informationen im Internet sind zentrale „Stellschrauben", die Einfluss auf die Wirksamkeit der Desinformation haben können. Durch gezielte Bildungsarbeit könnten diese bereits frühzeitig in den Blick genommen werden, um sowohl das Interesse an Politik und komplexen Zusammenhängen sowie Kompetenzen im Umgang mit Informationsmedien zu schulen (s. Kapitel 4, B.II.).

Besondere Aufmerksamkeit sollte auf die Ausbildung eines analytischen Denkvermögens gelegt werden, da Personen mit einer stärkeren Neigung zum analytischen Denken Falschinformationen besser erkennen können. Ein hohes Bedürfnis nach kognitiver Betätigung hängt positiv mit dem Erkennen von korrekten oder inkorrekten Informationen zusammen. Dies spricht ebenfalls dafür, die Neigung, sich mit komplexen Inhalten auseinanderzusetzen, bereits in der Schule zu fördern (s. Kapitel 4, A.III.).

2. Medienkompetenz

Medienkompetenzschulungen sind jedoch nicht nur für Kinder und Jugendliche, sondern auch für Erwachsene sinnvoll, die möglicherweise nicht so gut mit den digitalen Möglichkeiten der Informationserstellung vertraut sind. Zwischen dem Alter der Versuchsteilnehmenden und der Glaubwürdigkeitsbewertung von Falschinformationen besteht ein positiver Zusammenhang, der zeigt, dass die Desinformation von älteren Personen glaubwürdiger wahrgenommen wurde (s. Kapitel 4, A.III.).

3. Sensibilisierung

Weitere Schwerpunkte der Medienbildung im Hinblick auf Desinformation sollten eine grundlegende Kenntnis professioneller journalistischer Kriterien zur Auswahl und sprachlichen Aufbereitung von Nachrichten, die Entwicklung von Fähigkeiten zum Gebrauch einfacher Instrumente zur Überprüfung von Desinformation sowie eine Sensibilisierung für die Kommunikationsdynamiken beim Teilen von Informationen auf Online-Plattformen sein (s. Kapitel 2, F.II.).

II. Gesetzliche Feinjustierungen

1. Ausgeglichene Regelungen

Ein größerer Korrekturbedarf hinsichtlich der materiellen Rechtslage besteht nicht. Eine neue, pauschale rechtliche Regelung zu Desinformation neben den bereits vorhandenen Straftatbeständen erscheint ebenso wenig erforderlich wie praktisch realisierbar. Ein neuer Straftatbestand würde vor allem symbolische Bedeutung haben und in der Praxis wenig wirksam sein. Eine solche Vorschrift könnte sich sogar nachteilig auf die Motivation von Betreibern und Nutzenden von Online-Plattformen auswirken, selbstverantwortlich und freiwillig Maßnahmen zum Kampf gegen Desinformation zu ergreifen. Hierzu zählen auch ihre technischen Strukturen, die stetig fortentwickelt werden. Sie werden auch im Hinblick auf die Ermittlung und das Ranking anhand der Faktentreue von Websites Fortschritte erzielen. Soweit die Desinformation ehrverletzende, unwahre Tatsachenbehauptungen über Personen enthält, stehen bereits die Straftatbestände der Verleumdung und der üblen Nachrede zur Verfügung (s. Kapitel 5, A.).

2. Schutz der Meinungsfreiheit

Im Übrigen setzt die Meinungsfreiheit dem Handlungsspielraum des Gesetzgebers enge Grenzen. Nicht in den Schutzbereich der Meinungsfreiheit fallen lediglich solche Äußerungen, die reine Tatsachenbehauptungen darstellen, also dem Beweis zugänglich sind. Sie können einfacher gelöscht werden. Soweit die Äußerung jedoch „Elemente des Dafürhaltens“ aufweist, liegt eine Meinungsäußerung vor, die auch dann insgesamt als Meinungsäußerung

zu qualifizieren ist, wenn sie meinungsbezogene Tatsachenbehauptungen enthält. Im Zweifel ist von einer Meinungsäußerung auszugehen. Diese darf nur reguliert werden, wenn die Vorschriften dem Schutz der Jugend und dem Recht der persönlichen Ehre dienen oder wenn sie Themen betreffen, die nichts mit einer spezifischen Meinung zu tun haben. Es verbleibt ein weiter Bereich der Desinformation, der rechtlich nicht verboten ist. Eine freie Demokratie muss sie aushalten und sich ihr immer wieder im öffentlichen Meinungskampf stellen (s. Kapitel 5, A.II.).

3. Nachbesserungen des NetzDG

Im Rahmen des NetzDG sind Nachbesserungen notwendig, um den Schutz von Autoren zu erhöhen, deren Beiträge zu Unrecht gesperrt oder gelöscht worden sind. Bislang schreibt das NetzDG den Kommunikationsplattformen kein Verfahren vor, blockierte Nutzerinhalte erneut zu prüfen und unrechtmäßig entfernte Inhalte wiederherzustellen (s. Kapitel 5, D.II.).

4. Transparenz von Social Bots

Im Hinblick auf die Regulierung von Social Bots sollten zuvörderst die Online-Plattformen durch Selbstverpflichtungen, klare Allgemeine Geschäftsbedingungen und ihre Durchsetzung den Einsatz von Social Bots regulieren. Auch wenn die Verbreitung von Desinformation bisher (noch) nicht überwiegend durch Malicious Social Bots erfolgt, so bewirken sie doch Verstärkereffekte der Desinformation, die vermeintliche Mehrheitsverhältnisse vortäuschen und die Reichweite der Desinformation erheblich erweitern können. Ihr Einsatz ist häufig für die Rezipienten nicht erkennbar. Ob Selbstverpflichtungen der Online-Plattformen, insbesondere die im „EU-Praxiskodex Desinformation“ vorgesehenen, eher vage formulierten Maßnahmen zum Umgang mit Fake-Accounts und Online-Bots, künftig genügen werden, muss vor einem gesetzgeberischen Tätigwerden überprüft werden. Ein pauschales Verbot von Social Bots wäre unverhältnismäßig. Eine Kennzeichnungspflicht und Transparenzvorgaben für Social Bots – wie auch der Entwurf für einen neuen Medienstaatsvertrag vorschlägt – sind jedoch möglich und mit der Meinungsfreiheit vereinbar (s. Kapitel 5, C.V.).

III. Rechtsdurchsetzung verbessern

1. Strafverfolgung verbessern

Soweit Desinformation strafbar ist, sollte die Strafbarkeit der Äußerungen eine größere abschreckende Wirkung entfalten, als es bislang der Fall ist. Eine bessere Rechtsdurchsetzung bei strafbaren Äußerungen im Internet sowie die schnelle Klärung von Fällen und die Zuführung der Täter zu Strafverfahren ist erforderlich, um (potenzielle) Täter von strafbarer Desinformation und Hetze im Internet künftig abzuhalten und die Zahl dieser Straftaten nachhaltig zu reduzieren. Das NetzDG ist ein wichtiger Baustein zur Durchsetzung des geltenden Rechts. Auch wenn es nur bestimmte Formen strafbarer Beiträge erfasst, hilft es, die Zahl strafbarer Inhalte auf den Seiten der Online-Plattformen zu reduzieren. Außerdem hat es eine öffentliche, kontroverse Diskussion über den richtigen Umgang mit rechtswidrigen und gesellschaftlich schädlichen Inhalten auf Kommunikationsplattformen in Gang gebracht. Für besonders wichtig erachten wir zudem die Kooperation von Strafverfolgungsbehörden und Online-Plattformen. So müssen strafbare Beiträge nicht nur nicht mehr sichtbar sein, sondern die Inhaltersteller und -verbreiter auch konsequent der Strafverfolgung zugeführt werden. Dazu beitragen könnte die Einführung einer Pflicht der Anbieter bei Löschung oder Sperrung wegen Offizialdelikten diese an Strafverfolgungsbehörden zu melden. Daneben sind jedoch auch weitere Maßnahmen, die die faktischen Voraussetzungen der Strafverfolgung betreffen, erforderlich (s. Kapitel 5, C.II.).

2. Sorgfaltspflicht durchsetzen

Hinsichtlich der Regulierung von Telemedienanbietern mit journalistisch-redaktioneller Prägung sind schärfere Sanktions- und Aufsichtsmöglichkeiten bei Verstößen gegen die Wahrheits- und Sorgfaltspflichten empfehlenswert. Die Telemedien mit journalistisch-redaktionell gestalteten Angeboten (weite Auslegung) sind nach den Regelungen des Rundfunkstaatsvertrags zur Einhaltung der anerkannten journalistischen Grundsätze und zur Prüfung von Nachrichten mit der nach den Umständen gebotenen Sorgfalt auf Inhalt, Herkunft und Wahrheit verpflichtet (s. Kapitel 5, A.I.).

Soweit jedoch keine Rechte Dritter betroffen sind, ist die Anwendung des Telemedienrechts auf Anbieter von Websites, die bewusst oder grob fahrlässig Falschmeldungen verbreiten, kaum wirkungsvoll. Im Rundfunk-

staatsvertrag sind keine Aufsichtsmaßnahmen vorgesehen. Da die bewusste Verbreitung von Desinformationen (z. B. zur Erzielung von Klicks für höhere Werbeeinnahmen oder zur Unterstützung einer bestimmten politischen oder verschwörerischen Ausrichtung) einen besonders schweren Verstoß gegen die journalistischen Wahrheits- und Sorgfaltspflichten darstellt und die öffentliche Meinungsbildung beeinflussen kann, sollte es möglich sein, entsprechende Richtigstellungen, Entfernungen oder andere Maßnahmen zu erwirken. Entsprechend könnten die Befugnisse der Landesmedienanstalten erweitert werden (s. Kapitel 5, A.IV. und D.III.).

3. Freiwillige Selbstkontrolle

Mit dem Deutschen Presserat existiert zudem ein Organ der freiwilligen Selbstkontrolle. Auch wenn Politik und Gesetzgebung hier aus gutem Grunde keinen direkten Einfluss haben, könnte eine Diskussion mit den beteiligten Akteuren angestoßen werden, ob bezogen auf Desinformation Sanktionsmöglichkeiten entwickelt werden können, die über eine öffentliche Rüge hinausgehen, was jedoch einen Bruch des auf Freiwilligkeit beruhenden Selbstregulierungskonzeptes bedeuten würde (s. Kapitel 5, A.IV.).

E. Empfehlungen für Einrichtungen der Forschungsförderung in der EU und in Deutschland

I. Anwendungsorientierte, interdisziplinäre Forschung stärken

1. Interdisziplinäre Forschung

Die anwendungsorientierte Forschung zu Desinformationen muss intensiviert werden, um verantwortliche Akteure in die Lage zu versetzen, Gesellschaft und Wirtschaft vor der Zunahme schädlicher Desinformation wirksam zu schützen. Bezüglich Einzelaspekten von Desinformation liegen zwar Ergebnisse vor, allerdings sind diese durch disziplinäre Sichtweisen eingeschränkt und auch innerhalb von Disziplinen fragmentiert (s. Kapitel 2, F.IV.).

Ein Beispiel für die Notwendigkeit der Zusammenführung von Forschungsansätzen selbst innerhalb einer Disziplin ist die Erkennung von („bösartigen") Social Bots, die Desinformation vollautomatisch verteilen und verstärken. Hier gibt es eine Reihe von Mechanismen, die jeweils eine einzige statische

Eigenschaft oder ein Spezifikum des Verhaltens dieser Bots erkennen. Gebraucht werden aber Kombinationsansätze, die die Erkennungsrate messbar steigern und sogar Social Bots erkennen können, deren Existenz zum Beispiel mit adversarial machine learning verschleiert werden soll (s. Kapitel 4, B.V.).

Anwendungsorientierte Forschung kann beispielsweise Wege aufzeigen, wie mittels Technikeinsatz faktenbasierter Journalismus gestärkt werden kann. Exemplarisch hierfür steht ein Mechanismus zur nachprüfbaren Zitierweise: Durch kryptographische Mechanismen lassen sich die Unverändertheit (Integrität) und die Echtheit der angegebenen Quelle (Authentizität) eines Zitats in einem Text überprüfen (Kreutzer, Niederhagen, Shrishak & Fhom, 2019). Weitere Forschung ist notwendig, um verstehen zu können, ob dieses und andere Ergebnisse tatsächlich transferierbar sind.

Die Themenfelder für Angstnarrative von Desinformationskampagnen entwickeln sich im Laufe der Zeit sicherlich weiter bzw. verschieben sich. Um darüber gesicherte Aussagen machen zu können, ist ebenfalls interdisziplinäre Forschung notwendig. In den ersten drei Quartalen von 2019 zeigten beispielsweise Umfragen auf, dass das Themenfeld Ausländer/Migration/Flüchtlinge von wesentlich weniger Menschen in Deutschland als „wichtiges Problem" wahrgenommen wird, als in dem Zeitraum, der im Rahmen der DORIAN-Samples untersucht wurde (Forschungsgruppe Wahlen, 2019). Dem Themenfeld Umwelt und Energiewende wird in der Bevölkerung seit Anfang 2019 eine wachsende Bedeutung zugeschrieben (Forschungsgruppe Wahlen, 2019). In weitergehenden Forschungsaktivitäten könnte untersucht werden, ob ähnliche Angstnarrative wie zum Thema Migration und innere Sicherheit auch zu anderen Themen aufgebaut werden, beispielsweise in Bezug auf Dieselfahrverbote oder in Bezug auf eine angeblich drohende „Ökodiktatur". Ein Vergleich der Argumentationsstruktur in Fake News zu unterschiedlichen Themen könnte hier interessante und aktuelle Erkenntnisse für alle beteiligten Disziplinen liefern. So könnten diese helfen den Einsatz von maschinellem Lernen zur Erkennung von Desinformation weiterzuentwickeln.

2. Maschinelles Lernen stärken

In DORIAN wurde damit begonnen, maschinelles Lernen für die Erkennung von Desinformation zu erforschen. Die Zwischenergebnisse dieses Ansatzes sind sehr vielversprechend. Allerdings besteht noch erheblicher Forschungsbedarf, um dieses Werkzeug effizient und effektiv einsetzen zu können. Auch in anderen Bereichen der Technikforschung und in den anderen Disziplinen

von DORIAN werden die gewählten Forschungsansätze durch den Erfolg bestätigt. Die dort umgesetzten Pionierarbeiten müssen fortgesetzt werden, um ihr Nutzenpotenzial entfalten zu können (s. Kapitel 4, C.).

3. Gesamthafte Sicht ermöglichen

Ein technischer Ansatz allein würde aber keine empirisch fundierten Maßnahmenvorschläge liefern, wie gesellschaftliche Akteure Desinformation begegnen können. Die Sicht der Technik muss mindestens ergänzt werden durch die Beiträge der Journalistik, der Psychologie, des Rechts, der Sozialwissenschaften und der Politikwissenschaft. Für die nächsten fünf Jahre werden mindestens die beiden nachfolgenden Zielstellungen der Forschungsförderung zu Desinformation empfohlen:

II. Verbreitungswege und Verbreitungsgrad von Desinformation erforschen

Um das Phänomen Desinformation in Deutschland erfassen zu können, braucht es belastbare Informationen über die tatsächlichen Verbreitungswege auf Websites und Online-Plattformen (Social Networks). Hierzu müssen technische Mechanismen entwickelt werden, die dies quantitativ und qualitativ auf den verschiedenen Desinformationskanälen ermitteln können.

III. Wirkungsweise und Wirkmächtigkeit von Desinformation erforschen

Um Desinformation effektiv bekämpfen zu können, sind die Erforschung der Wirkungsweise, Wirkmächtigkeit und Wirkmaximierung notwendig – sowohl von Desinformation und in Kontrast dazu von faktenbasierten Informationen ohne Manipulationsabsicht.

Literaturverzeichnis zu Kapitel 6

Bader, K., Jansen, C., Johannes, P. C., Krämer, N., Kreutzer, M., Löber, L. I., Rinsdorf, L., Rösner, L. & Roßnagel, A. (2018). Desinformationen aufdecken und bekämpfen: Handlungsempfehlungen. Policy-Paper. Hrsg.: Fraunhofer-Institut für System- und Innovationsforschung ISI. Verfügbar unter https://www.forum-privatheit.de/wp-content/uploads/Policy-Paper-DORIAN-Desinformation-aufdecken-und-bekaempfen-1.pdf

Forschungsgruppe Wahlen (2019). Abruf am 23.9.2019 Verfügbar unter: http://www.forschungsgruppe.de/Umfragen/Politbarometer/Langzeitentwicklung_-_Themen_im_Ueberblick/Politik_II/

Gabielkov, M., Ramachandran, A., Chaintreau, A., & Legout, A. (2016). Social clicks: What and who gets read on Twitter?. ACM SIGMETRICS Performance Evaluation Review, 44(1), 179-192.

Kreutzer, M., Niederhagen, R., Shrishak, C. & Fhom, H. (2019). Quotable Signatures using Merkle Trees. INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik–Informatik für Gesellschaft.

Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect?. Political Communication, 35(2), 196-219.

Sängerlaub, A., Meier, M., & Rühl, W.-D. (2018). Fakten statt Fakes: Das Phänomen »Fake News«. Verursacher, Verbreitungswege und Wirkungen von Fake News im Bundestagswahlkampf 2017 (Abschlussbericht Projekt »Measuring Fake News«). Berlin. Verfügbar unter: https://www.stiftung-nv.de/sites/default/files/snv_fakten_statt_fakes.pdf

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck. org and FlackCheck. org. Journalism & Mass Communication Quarterly, 95(1), 49-75.

Autorenindex

Prof. Dr. Katarina Bader

Hochschule der Medien Stuttgart

Oren Halvani

Fraunhofer-Institut für Sichere Informationstechnologie

Wendy Freifrau Heereman von Zuydtwyck

Fraunhofer-Institut für Sichere Informationstechnologie

Michael Herfert

Fraunhofer-Institut für Sichere Informationstechnologie

Birte Högden

Universität Duisburg-Essen

Dr. Carolin Jansen

Hochschule der Medien Stuttgart

Paul Christopher Johannes

Universität Kassel

Prof. Dr. Nicole Krämer

Universität Duisburg-Essen

Dr. Michael Kreutzer

Fraunhofer-Institut für Sichere Informationstechnologie

Dr. Huajian Liu

Fraunhofer-Institut für Sichere Informationstechnologie

Lena Isabell Löber

Ass. iur., Universität Kassel

Judith Meinert

Universität Duisburg-Essen

Prof. Dr. Lars Rinsdorf

Hochschule der Medien Stuttgart

Prof. Dr. Alexander Roßnagel

Universität Kassel

Dr. Leonie Schaewitz

Universität Duisburg-Essen

Hervais-Clemence Simo Fhom

Fraunhofer-Institut für Sichere Informationstechnologie

Prof. Dr. Martin Steinebach

Fraunhofer-Institut für Sichere Informationstechnologie

Inna Vogel

Fraunhofer-Institut für Sichere Informationstechnologie

Ruben Wolf

Fraunhofer-Institut für Sichere Informationstechnologie

York Yannikos

Fraunhofer-Institut für Sichere Informationstechnologie

Dr. Sascha Zmudzinski

Fraunhofer-Institut für Sichere Informationstechnologie