

**GESELLSCHAFTLICHE
AUSWIRKUNGEN SYSTEMISCHER
RISIKEN**

**Demokratische Prozesse im Kontext von
Desinformationen**

Ann-Kathrin Watolla, Patrick Zerrer, Jan Rau, Lisa Merten,
Matthias C. Kettemann, Cornelius Puschmann

GESELLSCHAFTLICHE AUSWIRKUNGEN SYSTEMISCHER RISIKEN

Demokratische Prozesse im Kontext von Desinformationen

Ann-Kathrin Watolla

Alexander von Humboldt-Institut für Internet und Gesellschaft, Berlin

Patrick Zerrer

Universität Bremen

Jan Rau

Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, Hamburg

Lisa Merten

Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, Hamburg

Matthias C. Kettemann

Alexander von Humboldt-Institut für Internet und Gesellschaft, Berlin

Cornelius Puschmann

Universität Bremen

Lektorat: Paul Stoop

Berlin/Bremen/Hamburg, März 2025

ZUSAMMENFASSUNG

Mit der vorliegenden Studie wird ein kritischer Blick auf die gesellschaftlichen Auswirkungen systemischer Risiken auf Online-Plattformen und Online-Suchmaschinen im Kontext demokratischer Prozesse geworfen. Hierbei stehen insbesondere Desinformationen im Fokus. Nach einer Verortung der Studie in einem größeren Kontext wird im ersten Teil nachgezeichnet, wie sich systemische Risiken auf digitalen Diensten realisieren und welche gesellschaftlichen Auswirkungen dies mit sich bringt. Ausgehend vom Begriff der „systemischen Risiken“ im Digital Services Act (DSA) werden diese verstanden als Gefährdungen, die sich aus den gegenseitigen Abhängigkeiten von Institutionen und Akteur:innen in der gesellschaftlichen Kommunikation sowie aufgrund der zentralen Funktion digitaler Dienste als Orte der kollektiven und individuellen Meinungsbildung ergeben. Eine einheitliche Definition von Desinformation gibt es nicht. In dieser Studie werden darunter verstanden: nachweislich falsche oder irreführende Informationen, die intentional für politische oder ökonomische Zwecke gestreut werden und das Potenzial für gesellschaftlichen Schaden haben. Dabei werden nicht nur einzelne Inhalte betrachtet, die falsche oder irreführende Informationen beinhalten, sondern vielmehr eine Vielzahl solcher Inhalte, die Teil größerer Narrative sind. Desinformationen in diesem Sinne sind kein isoliertes Phänomen, sondern sind im größeren Kontext der Prozesse von Meinungsbildung durch Online-Kommunikation zu betrachten.

Die Verbreitung von Desinformationen erfolgt durch verschiedene Akteur:innen, darunter ausländische staatliche Akteur:innen, politische Parteien, Aktivist:innen und Influencer:innen. Die zugrundeliegenden Motive für die Verbreitung von Desinformationen reichen von politischer Einflussnahme bis hin zu kommerziellen Interessen in der digitalen Aufmerksamkeitsökonomie, die darauf beruht, dass Aufmerksamkeit ein knappes Gut ist, das sich in Geld oder politischen Einfluss umwandeln lässt. Daher wird immer mehr Aufwand betrieben, um sie gezielt zu erzeugen, insbesondere mit negativ-emotionalisierten und Gruppenidentitäts-basierten Inhalten. Die Verbreitung von Desinformationen erfolgt auf verschiedenen Wegen. Dazu gehören koordiniertes inauthentisches Verhalten für die massenhafte Verbreitung von Desinformationen, Priorisierung der Sichtbarkeit von polarisierenden und emotionalisierenden Inhalten durch Empfehlungsalgorithmen, Manipulation von Suchmaschinen zur Priorisierung irreführender Inhalte sowie die Produktion und Verbreitung von Desinformationen mithilfe von Künstlicher Intelligenz (KI) – und insbesondere generativer KI –, z.B. von Deepfakes.

Daran anschließend behandelt der zweite Teil der Studie die Auswirkungen systemischer Risiken auf Gesellschaft, wobei Desinformationen im Kontext demokratischer Prozesse im Fokus stehen. In der Studie wird dargelegt, dass eindeutige Kausalitäten nur schwer nachgewiesen werden können, Grundlegend lässt sich jedoch eine konfliktorientierte Verschiebung des öffentlichen Diskurses durch digitale Medien als Bedrohung für eine wichtige Funktion politischer Öffentlichkeit identifizieren: kollektives Handeln zu ermöglichen. Diese

Verschiebung begünstigt einen diskursiven Stil, der durch Gruppenzugehörigkeiten, Konflikte und moralisierende Sprache geprägt ist. Desinformationen profitieren von diesen Mechanismen emotionaler Empörung und tragen gleichzeitig zur weiteren Polarisierung bei. Diese Polarisierung kann ideologisch, affektiv und themenbezogen sein. Die verschiedenen Formen der Polarisierung können sich gegenseitig verstärken und zu einer destruktiven Polarisierung führen, die eine Gefahr für demokratische Prozesse und den gesellschaftlichen Zusammenhalt darstellt. Da die digitale Transformation den öffentlichen Diskurs und die Interaktion zwischen Bürger:innen, politischen Akteur:innen und Institutionen grundlegend verändert hat, ergeben sich zudem neue Herausforderungen für das Vertrauen in politische Prozesse. Vertrauen wird als zentrale Grundlage für das Funktionieren demokratischer Gesellschaften identifiziert. Die oben genannte Diskursverschiebung geht mit einem Vertrauensverlust in demokratische Prozesse und Institutionen einher. Der Verlust dieses Vertrauens kann die Handlungsfähigkeit demokratischer Institutionen beeinträchtigen und die Bereitschaft zu Kompromissen verringern.

Desinformationen sind in diese Entwicklungen eingebettet, wobei jedoch zwischen der Existenz von Desinformationen auf Online-Plattformen und -Suchmaschinen, der Exposition durch Nutzende sowie den daraus resultierenden Effekten z. B. auf Wahlverhalten unterschieden werden muss. Während die Existenz von großangelegten Desinformationskampagnen – wie beispielsweise der russischen Doppelgänger-Kampagne – klar nachgewiesen werden kann, gibt es bis dato kaum Studien zur Messung der Exposition von Desinformationen und zu den daraus resultierenden Effekten auf das Wahlverhalten. Insofern bedarf es einer kritischen Betrachtung, inwiefern die reine Existenz von Desinformationen in einen kausalen Zusammenhang mit z. B. Wahlergebnissen gestellt werden kann. Dies ist umso relevanter, als empirische Studien zeigen, dass Desinformationen in der öffentlichen Debatte als großes gesellschaftliches Problem wahrgenommen werden. Diese Wahrnehmung von – tatsächlicher oder vermeintlicher – Desinformation wird als eine Gefahr für Demokratie und gesellschaftlichen Zusammenhalt betrachtet und wirkt sich zudem negativ auf das institutionelle Vertrauen aus.

In Bezug auf die Bewertung von Desinformationsgefahren im Kontext von KI-Systemen durch Nutzende zeigt sich ferner, dass die Mehrheit der Bevölkerung KI-Anwendungen wie ChatGPT skeptisch gegenübersteht und sich der Gefahren des Missbrauchs bewusst ist. Die Fähigkeiten zur Erkennung von Deepfakes in der Bevölkerung sind jedoch begrenzt. So zeigen Studien, dass etwa 40 % der Teilnehmenden Deepfakes nicht als manipuliertes Material identifizieren konnten, viele dabei jedoch ihre Fähigkeiten zur korrekten Identifikation überschätzen. Der bewusste Kontakt mit Deepfakes, etwa durch Kennzeichnungen, scheint die Unsicherheit der Rezipierenden zudem zu verstärken und das allgemeine Medienvertrauen zu mindern. Diese Erkenntnisse unterstreichen die Komplexität der Herausforderungen, die Desinformationen sowie KI-generierte Inhalte für die Gesellschaft und das Vertrauen in Medien und Institutionen darstellen. Hinzu kommt, dass wir aktuell noch zu wenig darüber wissen, wie viele Personen und wie oft tatsächlich mit Desinformationen konfrontiert sind (Exposition), oder ob diese Wahrnehmung davon beeinflusst wird, wie Desinformationen in der öffentlichen Debatte dargestellt werden.

Abschließend diskutiert die Studie Möglichkeiten zur Begrenzung der Auswirkungen von Desinformationen, wobei sie die Bedeutung von Medienkompetenz und kritischem Denken hervorhebt. Sie weist jedoch auch auf die Herausforderungen hin, die sich aus der zunehmenden Qualität und Personalisierung von KI-generierten Inhalten ergeben. Abschließend betont die Studie die Notwendigkeit weiterer Forschung, um die komplexen Zusammenhänge zwischen digitalen Medien, Desinformationen und gesellschaftlichen Veränderungsprozessen besser zu verstehen. Sie unterstreicht die Bedeutung eines ganzheitlichen Ansatzes, der technologische, soziale und politische Aspekte berücksichtigt, um den Herausforderungen von Desinformationen in demokratischen Gesellschaften zu begegnen.

EXECUTIVE SUMMARY

This study takes a critical look at the social impact of systemic risks on online platforms and online search engines in the context of democratic processes. The focus here is on disinformation in particular. After locating the study in a wider context, the first part traces how systemic risks are realised on digital services and what social effects this entails. Based on the concept of ‘systemic risks’ in the Digital Services Act (DSA), these are understood as hazards that arise from the mutual dependencies of institutions and actors in social communication as well as from the central function of digital services as places of collective and individual opinion-forming. There is no standardised definition of disinformation. In this study, it is understood to mean: demonstrably false or misleading information that is intentionally disseminated for political or economic purposes and has the potential to cause social harm. It is not just individual pieces of content that contain false or misleading information that are considered, but rather a variety of such content that are part of larger narratives. Disinformation in this sense is not an isolated phenomenon, but must be viewed in the wider context of the processes of opinion formation through online communication.

Disinformation is disseminated by various actors, including foreign state actors, political parties, activists and influencers. The underlying motives for spreading disinformation range from political influence to commercial interests in the digital attention economy, which is based on the fact that attention is a scarce commodity that can be converted into money or political influence. Therefore, more and more effort is being put into generating it in a targeted manner, especially with negative emotionalised and group identity-based content. Disinformation is spread in a variety of ways. These include coordinated inauthentic behaviour for the mass dissemination of disinformation, prioritising the visibility of polarising and emotionalising content through recommendation algorithms, manipulation of search engines to prioritise misleading content, and the production and dissemination of disinformation using artificial intelligence (AI) – and generative AI in particular – such as deepfakes.

The second part of the study then deals with the effects of systemic risks on society, focussing on disinformation in the context of democratic processes. The study demonstrates that clear causalities are difficult to prove, but fundamentally, a conflict-orientated shift in public discourse through digital media can be identified as a threat to an important function of the political public sphere: to enable collective action. This shift favours a discursive style that is characterised by group affiliations, conflicts and moralising language. Disinformation benefits from these mechanisms of emotional outrage and at the same time contributes to further polarisation. This polarisation can be ideological, affective and issue-related. The various forms of polarisation can reinforce each other and lead to destructive polarisation that poses a threat to democratic processes and social cohesion. As the digital transformation has fundamentally changed public discourse and the interaction between citizens, political actors and institutions, new challenges for trust in political processes are also emerging. Trust is identified as the

central basis for the functioning of democratic societies. The above-mentioned shift in discourse is accompanied by a loss of trust in democratic processes and institutions. The loss of this trust can impair the ability of democratic institutions to act and reduce the willingness to compromise.

Disinformation is embedded in these developments, although a distinction must be made between the existence of disinformation on online platforms and search engines, exposure by users and the resulting effects, e.g. on voting behaviour. While the existence of large-scale disinformation campaigns - such as the Russian doppelgänger campaign - can be clearly proven, there have been hardly any studies to date on measuring exposure to disinformation and the resulting effects on voting behaviour. In this respect, the extent to which the mere existence of disinformation can be placed in a causal relationship with election results, for example, needs to be critically analysed. This is all the more relevant as empirical studies show that disinformation is perceived as a major social problem in the public debate. This perception of - actual or perceived - disinformation is seen as a threat to democracy and social cohesion and also has a negative impact on institutional trust.

With regard to the assessment of disinformation risks in the context of AI systems by users, it also shows that the majority of the population is sceptical about AI applications such as ChatGPT and is aware of the dangers of misuse. However, the population's ability to recognise deepfakes is limited. Studies show that around 40% of participants were unable to identify deepfakes as manipulated material, but many overestimate their ability to correctly identify them. Deliberate contact with deepfakes, for example through labelling, also appears to increase the uncertainty of recipients and reduce general media trust. These findings emphasise the complexity of the challenges that disinformation and AI-generated content pose for society and trust in the media and institutions. In addition, we currently know too little about how many people are actually confronted with disinformation and how often (exposure), or whether this perception is influenced by how disinformation is presented in the public debate.

Finally, the study discusses ways to limit the impact of disinformation, emphasising the importance of media literacy and critical thinking. However, it also points to the challenges posed by the increasing quality and personalisation of AI-generated content. In conclusion, the study emphasises the need for further research to better understand the complex relationships between digital media, disinformation and social change processes. It emphasises the importance of a holistic approach that takes technological, social and political aspects into account in order to meet the challenges of disinformation in democratic societies.

INHALTSVERZEICHNIS

ZUSAMMENFASSUNG	3
EXECUTIVE SUMMARY	6
1 EINLEITUNG	9
2 METHODISCHES VORGEHEN	11
3 VORARBEITEN FÜR DAS SCOPING REVIEW	13
3.1 Zum Begriff der systemischen Risiken	13
3.2 Definitorische Annäherung an den Begriff der Desinformationen	15
3.3 Die digitale Transformation politischer Öffentlichkeit	17
3.4 Echokammern und Filterblasen	20
4 REALISIERUNG SYSTEMISCHER RISIKEN AUF DIGITALEN DIENSTEN	22
4.1 Hauptakteur:innen bei der Verbreitung von Desinformationen und ihre Motive	22
4.2 Technische Verbreitungsmöglichkeiten von Desinformationen in Online-Suchmaschinen und auf Online-Plattformen	24
4.2.1 Verbreitungsstrategien in Online-Suchmaschinen	24
4.2.2 Verbreitungsstrategien auf Online-Plattformen	25
4.3 Die Rolle von Empfehlungssystemen auf Online-Plattformen und in Online-Suchmaschinen bei der Verbreitung von Desinformationen	26
4.4 Bewertung des Gefährdungspotentials durch KI und durch diese generierten Inhalte	28
5 AUSWIRKUNGEN SYSTEMISCHER RISIKEN AUF DIE GESELLSCHAFT	33
5.1 Soziale und politischen Auswirkungen von Desinformationskampagnen und anderen Methoden der Beeinflussung	33
5.2 Wahrnehmung und Bewertung der Gefahren von Desinformationskampagnen und anderen Methoden der Beeinflussung	42
5.3 Feststellung und Quantifizierung der Auswirkungen	46
5.4 (Mit-)Ursächlichkeit von Desinformationen für die genannten Auswirkungen	56
5.5 Beobachtung neuester Entwicklungen	59
5.6 Begrenzung der Auswirkungen und Handlungsspielräume der Plattformen	62
6 FAZIT UND AUSBLICK	68
LITERATURVERZEICHNIS	73

1 EINLEITUNG

Mit Annahme des Digital Services Act (DSA) wurde 2022 eine neue Ära der Plattformregulierung eingeläutet. Zusammen mit dem Digital Markets Act, dem Data Act, dem Data Governance Act, dem European Media Freedom Act und dem AI Act erhöht die Europäische Union (EU) substantiell die Regelungsdichte in kommunikativen Online-Ökosystemen. Mit der Neuregulierung soll ein sicheres digitales Umfeld in der EU geschaffen werden, in dem die Grundrechte von Nutzenden geschützt sowie Innovation und Wettbewerb gefördert werden. Dabei zielt insbesondere der DSA auf den Schutz vor Verbreitung rechtswidriger Inhalte. Die Neuregulierung will Einfluss auf aktuelle Transformationsprozesse, politische Diskurse und Öffentlichkeiten für Politik und Gesellschaft der Union nehmen und ist gleichzeitig Folge dieser.

Ursprünglich als Katalysator für demokratische Teilhabe befürwortet, ist der öffentliche Diskurs im Internet und auf sozialen Diensten in den letzten zehn Jahren zunehmend als Problemfall angesehen worden. Digitale Dienste werden immer mehr im Kontext von Extremismus, Polarisierung und der Verbreitung von Desinformationen in WEIRD-Gesellschaften (Western, Educated, Industrialized, Rich, and Democratic) diskutiert (Lorenz-Spreen et al., 2023). Forscher:innen warnen jedoch davor, diese Entwicklungen ausschließlich dem Aufstieg digitaler Medien zuzuschreiben. Die Polarisierung der amerikanischen Öffentlichkeit etwa begann wohl schon mit Reagans anti-gouvernementaler Wende und wird von klassischen Medien wie dem linearen Fernsehen und kleineren Radiosendern substantiell verschärft. Auch strukturelle gesellschaftliche Umwälzungen, Konflikte und Krisen seien in diesem Kontext zu beachten. Gleichzeitig spielten digitale Medien eine wichtige Rolle bei der Entstehung und Verschärfung von Konflikten und Krisen, und trügen dazu bei, bestimmte Konflikte und Spannungen zu verstärken (Lorenz-Spreen et al., 2023; Schroeder, 2018).

In diese digitalen Transformationsprozesse greift der DSA als komplexes, „hybrides“ Gefüge von Steuerungsfaktoren ein (Schulz & Ollig, 2023). Der DSA zielt darauf ab, eine „sichere, vorhersehbare und vertrauenswürdige Online-Umgebung“ zu schaffen, die die „in der Charta verankerten Grundrechte“ wirksam schützt (Art. 1 Abs. 1 DSA). Die Regulierung befasst sich dafür mit einer breiten Palette von Themen und verfolgt einen gestaffelten Ansatz: Die strengsten Anforderungen gelten für sehr große Online-Plattformen (Very Large Online Platforms, VLOPs) und sehr große Online-Suchmaschinen (Very Large Search Engines, VLOSEs), die mehr als 45 Millionen Nutzende innerhalb der EU bedienen. Im Rahmen des risikobasierten Regulierungsansatzes des DSA (Efroni, 2021; Hofmann & Raue, 2023; Ederer, 2023; Roth-Isigkeit, 2024) müssen Anbietende von VLOPs und VLOSEs systemischen Risiken eigenständig begegnen (Art. 34, 35 DSA). Konkret verlangt Art. 34 DSA, dass VLOP- und VLOSE-Anbietende „sorgfältig alle systemischen Risiken in der Union identifizieren, analysieren und bewerten“, die sich aus ihren Diensten ergeben, und zwar mindestens einmal jährlich. Anschließend schreibt Art. 35 DSA vor, dass VLOP- und VLOSE-Anbietende „angemessene, verhältnismäßige und wirksame Minderungsmaßnahmen“ umsetzen müssen, die auf diese identifizierten Risiken

zugeschnitten sind. Dabei müssen die Anbietenden „besondere Berücksichtigung“ der potenziellen Auswirkungen auf die Grundrechte wahren lassen. Systemische Risiken sind im DSA selbst nur vage beschrieben. Es obliegt den Anbietenden digitaler Dienste, systemische Risiken zu konkretisieren und zu priorisieren (Griffin, 2023). Dies lässt vermuten, dass die Kategorien der in Art. 34 DSA genannten systemischen Risiken im Laufe der Zeit präzisiert und ausdifferenziert werden müssen, was wiederum die Bewertungsverfahren und Minderungsmaßnahmen beeinflussen wird.

Die vorliegende Studie stellt die potenziellen Auswirkungen systemischer Risiken auf Grundrechte anhand des Einflusses von Desinformationen auf demokratische Prozesse, insbesondere Wahlen, in den Mittelpunkt. Dieser thematische Zuschnitt ermöglicht es, Leerstellen in der aktuellen Forschung zu identifizieren und den Zusammenhang zwischen systemischen Risiken und gesellschaftlichen Auswirkungen genauer zu verstehen. Kernfrage ist dabei: Wie manifestieren sich Desinformationen auf sozialen Plattformen und welchen Einfluss nehmen sie auf Gesellschaft und demokratische Prozesse. Ein Scoping Review (siehe Kapitel 3) soll dafür einen breiten Überblick über theoretisierende und empirische Forschungserkenntnisse unterschiedlicher Disziplinen liefern. Den fachlichen Ausgangspunkt bilden dafür grundlegende Vorarbeiten in vier Bereichen: der Begriff der systemischen Risiken, die Problematisierung des Begriffs der Desinformationen, die digitale Transformation politischer Öffentlichkeit sowie der Diskurs über Echokammern und Filterblasen (Kapitel 3). Auf dieser Basis werden zunächst die Mechanismen und Prozesse von Desinformationsverbreitung in sozialen Medien nachgezeichnet (Kapitel 4), bevor diese in Hinblick auf ihre gesellschaftlichen Auswirkungen im Kontext demokratischer Prozesse betrachtet werden (Kapitel 5). In einem abschließenden Fazit werden zentrale Erkenntnisse der Studie konsolidiert und ein Ausblick auf offene Forschungsfragen gegeben (Kapitel 6).

Beim Bemühen um einen breiten Überblick über das heterogene Themenfeld ist zu berücksichtigen, dass ausschließlich öffentlich zugängliche Daten berücksichtigt werden können und dadurch ggfs. vertrauliche Erkenntnisse, z. B. von Geheimdiensten oder anderen staatlichen Stellen, nicht berücksichtigt werden können. Hinzu kommt, dass es aufgrund fehlender Transparenz über das methodische Vorgehen in einzelnen Studien nicht immer möglich ist, den Erkenntnisgewinn methodisch zu rekonstruieren und somit im Rahmen der Studie zu kontextualisieren.

2 METHODISCHES VORGEHEN

Diese dezidiert interdisziplinäre Studie nutzt Erkenntnisse aus den Politikwissenschaften, der Soziologie, der Kommunikationswissenschaft, der Psychologie, der Informatik, der Informationswissenschaft und der Human Computer Interaction-Forschung. Dieser Ansatz birgt große Chancen, ist aber auch eine Herausforderung, da den verschiedenen Disziplinen unterschiedliche Perspektiven zugrunde liegen. Auch muss bedacht werden, dass Wirkungsvermutungen in diesem komplexen Forschungsfeld nur schwer methodisch solide verifiziert werden können. Deshalb werden Scoping Reviews genutzt, die als „mapping,‘ a process of summarizing a range of evidence in order to convey the breadth and depth of a field“ (Levac et al., 2010, S. 1) erste Erkenntnisse über komplexe Themenbereiche liefern können. Um dem höchst dynamischen und interdisziplinären Forschungskontext der Studie Rechnung zu tragen, wird dem Scoping Review eine Vielfalt an Daten zugrunde gelegt: Neben peer-reviewter Forschungsliteratur wird auch „graue“ Literatur¹ verwendet, um auch aktuelle Forschungserkenntnisse berücksichtigen zu können. Diese Quellen umfassen theoretisierende wie auch empirische Forschungsarbeiten. Darüber hinaus werden im Scoping Review auch Berichte² von Anbietenden von VLOPs und VLOSEs genutzt, um so weitere Einblicke zu erhalten, wie sich systemische Risiken und insbesondere Desinformationen im Kontext demokratischer Prozesse auf VLOPs und VLOSEs realisieren und welche Minderungsmaßnahmen hierfür eingesetzt werden.

Dem Scoping Review wurden mit Blick auf den Forschungskontext drei Aufnahmekriterien zugrunde gelegt:

1. Da hier der DSA und die damit verbundenen regulatorischen Neuerungen im Mittelpunkt stehen, wird im Scoping Review vor allem Literatur berücksichtigt, die seit 2022 erschienen ist. Diese Grenze lässt eine Betrachtung des Forschungsfeldes in den Forschungsstudien unter den Bedingungen des DSA vermuten, selbst wenn die Wirkungen des DSA weder 2022 noch heute kaum messbar sind. Es werden aber auch ältere Quellen herangezogen, in denen grundlegende Entwicklungen und Erkenntnisstände nachgezeichnet werden.
2. Da der deutsche bzw. europäische Diskursraum im Mittelpunkt stehen soll, werden hier vor allem solche Forschungsstudien über digitale Dienste einbezogen, die in diesem

¹ Mit grauer Literatur werden Veröffentlichungen bezeichnet, die nicht der Kontrolle des Peer Reviews oder des Verlagswesens unterliegen, sondern von Behörden, Nichtregierungsorganisationen, Forschungseinrichtungen etc. eigenständig veröffentlicht werden.

² Dazu gehören sowohl die Berichte der Trust & Security Teams über Angriffe und Bedrohungen als auch die entsprechend Art. 34 DSA erstmals im November 2024 bereitgestellten Risikobewertungen.

geopolitischen Raum am häufigsten genutzt werden³. Für das Nachzeichnen globaler Phänomene und Entwicklungen werden auch Forschungsstudien genutzt, die über diesen Raum hinausgehen.

3. Gleiches gilt für die Nutzung von Studien, die systemische Risiken analysieren. Auch hier werden vorrangig solche berücksichtigt, die sich auf den deutschen bzw. europäischen Diskursraum beziehen. Internationale Forschungsstudien sollen an dieser Stelle lediglich zur Kontextualisierung der Entwicklungen in Deutschland bzw. Europa dienen.

Auf Basis dieser Kriterien wurde die herangezogene Literatur in Bezug auf etablierte Gütekriterien der qualitativen Sozialforschung bewertet und entsprechend für das Scoping Review ausgewählt (Strübing, 2018). Zentrales Anliegen dabei ist auch, die Heterogenität des Forschungskontexts angemessen widerzuspiegeln.

³ Für die am häufigsten in Deutschland genutzten Dienste liefert bspw. die ARD/ZDF-Medienstudie aussagekräftige Erkenntnisse: <https://www.ard-media.de/media-perspektiven/studien/ard/zdf-medienstudie>

3 VORARBEITEN FÜR DAS SCOPING REVIEW

Im Folgenden gehen wir schrittweise vor: Zunächst erfolgt eine Diskussion des Begriffs der „systemischen Risiken“, vor allem angesichts unscharfer Begriffsverwendungen im DSA. Dies dient der Bildung eines Grundverständnisses des hier behandelten Themas. Der zweite Schritt ist eine definitorische Annäherung an den Begriff der Desinformationen. Anschließend werden grundlegende Entwicklungen der digitalen Transformation politischer Öffentlichkeit dargelegt, bevor in einem letzten Schritt die vielfach diskutierten Echokammern und Filterblasen kritisch beleuchtet werden.

3.1 Zum Begriff der systemischen Risiken

Den systemischen Risiken im DSA liegt die Beobachtung zugrunde, dass „die Art und Weise, in der sehr große Online-Plattformen und sehr große Online-Suchmaschinen genutzt werden können, [...] großen Einfluss auf die Online-Sicherheit, die öffentliche Meinungsbildung und den öffentlichen Diskurs sowie den Online-Handel“ hat (ErwGr 79 DSA). Dafür werden vier Kategorien solcher systemischer Risiken benannt (ErwGr 80 DSA), allerdings ohne weitergehende Präzisierung des Begriffs der systemischen Risiken. Daher muss der Begriff an dieser Stelle umrissen werden. In der Forschung wird festgehalten, dass der Begriff der systemischen Risiken ursprünglich dem Finanzsektor entstammt. Deshalb bedarf die Übertragung auf Plattformregulierung einer Neuinterpretation der Begriffsbedeutung. Wie eine tiefgehende Betrachtung der sektorspezifischen Kontexte systemischer Risiken zeigt (Broughton Micova & Calef, 2023), bezieht sich diese zum einen darauf, was als systemisch verstanden wird und zum anderen darauf, wie sich Risiken innerhalb dieses Systems manifestieren. Im Finanzsektor ist die systemische Relevanz von Institutionen vor allem von zwei Faktoren geprägt: zum einen von den starken Interdependenzen von Institutionen miteinander bzw. Abhängigkeiten voneinander und zum anderen von der damit zusammenhängenden Größe und Marktbedeutung der jeweiligen Institutionen. In Bezug auf die Verflechtungen von Institutionen wird demnach von systemischen Risiken gesprochen, wenn das Risiko eines „systemischen Ausfalls“ besteht, bei dem der Zusammenbruch eines großen Teils des Finanzsystems, ausgehend von einer oder mehreren Finanzinstitutionen, das gesamte System destabilisieren kann (Pacelli et al., 2024). Dazu gehört auch, dass zahlreiche Finanzinstitutionen gleichzeitig Verluste erleiden (Acharya et al., 2017; Silva et al., 2017). Bei der Übertragung auf Plattformregulierung zeigt sich, dass „das System“ im Kontext des DSA aufgrund der vielfältigen Akteur:innen und deren Verflechtungen auf unterschiedlichen Ebenen nicht so einfach vorzunehmen ist wie im Finanzsektor, wo das Finanzsystem klar abgegrenzt werden kann.

Auch sind – auf den Kontext des DSA übertragen – gleichzeitige Verluste mehrerer digitaler Dienste kein realistisches Szenario, da die unterschiedlichen digitalen Dienste unabhängig voneinander agieren. Dennoch lassen sich in verschiedener Hinsicht Interdependenzen zwischen digitalen Diensten und Staaten, politischen Akteur:innen, Unternehmen und weiteren Stakeholdern verzeichnen, die Gefahren für Rechtsgüter darstellen. Insofern bietet es sich an, die Verflechtungen von Institutionen, wie sie im Kontext des Finanzsektors erkennbar sind, auf den Kontext des DSA zu übertragen, auch wenn sich diese nicht im Sinne eines potenziellen systemischen Ausfalls manifestieren, sondern eher in ihren gegenseitigen Abhängigkeiten. So spielen Staaten eine wichtige Rolle bei der Einrichtung und Garantie der Strukturbedingungen kommunikativer Prozesse und der privaten und öffentlichen Kommunikation, selbst in hybriden Kommunikationsräumen. Besonders die zunehmende Rolle von digitalen Diensten in politischer Kommunikation sowie die Verknüpfung von Akteur:innen, Infrastrukturen oder Medieninhaber:innen sind hier hervorzuheben, wie auch Broughton Micova und Calef (2023) in ihrer Analyse der Neuinterpretation des Begriffs der systemischen Risiken im Kontext von Plattformregulierung aufzeigen. Aktuelle Beispiele für die Verbindungen von Plattformen und Politik finden sich beispielsweise in der Rolle von X im Wahlkampf der AfD⁴ im Vorfeld der deutschen Bundestagswahlen 2025 (Nenno & Lorenz-Spreen, 2025) oder im amerikanischen Präsidentschaftswahlkampf 2024 (Duffy, 2024), von Infrastrukturen und Plattformen im Zusammenwirken vom Satellitennetzwerk Starlink und X gegen die Entscheidung des Obersten Gerichtshof in Brasilien („Blockade in Brasilien“, 2024) sowie von Plattformen und Nachrichtenagenturen in der Drohung von Google, Investitionen ihrer Google News Initiative in die Nachrichtenbranche zu stoppen (Krempf, 2024).

Zudem sind digitale Dienste insofern systemisch, als sie für gesellschaftliche Diskurse und Meinungsbildung als zentrale Plattformen im Informationsökosystem fungieren und ihnen durch ihre Reichweite bestimmte gesellschaftliche Funktionen zugeschrieben werden. Der DSA rechnet zu den sehr großen Online-Plattformen und -Suchmaschinen solche, die eine durchschnittliche monatliche Zahl von mindestens 45 Millionen aktiven Nutzenden in Europa (Art. 33 Abs. 1 DSA) zählen. Das entspricht etwa 10 Prozent der EU-Bevölkerung. Mit dieser Festschreibung gehen strengere Regeln einher, wie z. B. Transparenzpflichten⁵ (Art. 15, 24, 27, 39, 42 DSA) und das Erstellen von Risikobewertungen (Art. 34, 35 DSA). Die systemische Natur digitaler Dienste ergibt sich demnach aus der Zahl betroffener Nutzenden und aus der Funktion sehr großer Online-Plattformen und -Suchmaschinen als öffentliche Orte der individuellen und kollektiven Meinungsbildung und dem damit einhergehenden Einfluss auf gesellschaftliche Diskurse.

Unter „Risiken“ versteht der DSA die tatsächliche oder potenzielle Gefährdung von Grundrechten durch digitale Dienste, wie sie in der Charta der Grundrechte der Europäischen Union (2012) verankert sind. Diese Gefährdung kann sowohl durch die Funktionsweisen der

⁴ Die öffentliche Unterstützung Elon Musks für die AfD allgemein und Alice Weidel im Speziellen führt zu einer erhöhten Reichweite dieser auf X.

⁵ Die Transparenzpflichten gelten jedoch nicht ausschließlich für VLOPs und VLOSEs.

digitalen Dienste als auch durch die Nutzungsweisen aus diesen entstehen (Broughton Micova & Calef, 2023). Konkret verweist der DSA auf folgende systemische Risiken (Art. 34 Abs. 1 a-d DSA):

- die Verbreitung rechtswidriger Inhalte
- tatsächliche oder absehbar nachteilige Auswirkungen auf die Ausübung der Grundrechte
- tatsächlichen oder absehbare nachteilige Auswirkungen auf die gesellschaftliche Debatte und auf Wahlprozesse und die öffentliche Sicherheit
- tatsächliche oder absehbare nachteilige Auswirkungen in Bezug auf geschlechtsspezifische Gewalt, den Schutz der öffentlichen Gesundheit und von Minderjährigen sowie schwerwiegende nachteilige Folgen für das körperliche und geistige Wohlbefinden einer Person.

Bei diesen Kategorien handelt es sich um Beispiele der systemischen Risiken, die von der Nutzung von Online-Plattformen und -Suchmaschinen ausgehen können. Allgemein werden systemische Risiken in dieser Studie als jene Gefahren verstanden, die sich aus den Verflechtungen von Institutionen und Akteur:innen in der gesellschaftlichen Kommunikation sowie durch die zentrale Funktion von digitalen Diensten als öffentliche Orte individueller und kollektiver Meinungsbildung ergeben.

3.2 Definitorische Annäherung an den Begriff der Desinformationen

Der Begriff der Desinformationen hat in den letzten Jahren zunehmend Einzug in den öffentlichen Diskurs erhalten, wird er doch von Forschenden, Politiker:innen, Medienproduzent:innen und -verteiler:innen gleichermaßen verwendet. Die Allgegenwart des Begriffs erfordert daher eine definitorische Annäherung. Dabei muss berücksichtigt werden, dass es keine einheitliche Begriffsdefinition gibt (Dreyer et al., 2021), die für die vorliegende Studie zugrunde gelegt werden kann. Zwar wurde bereits vor einigen Jahren der Unterschied zwischen Desinformationen, Misinformationen und Falschinformationen aufgezeigt (Jack, 2017; Wardle & Derakhshan, 2017), es zeigt sich allerdings sowohl in den Regelwerken der EU als auch in denen der sehr großen Online-Plattformen und -Suchmaschinen, dass mit einem Begriff nicht immer dasselbe gemeint (Bleyer-Simon & Reviglio, 2024). Deshalb ist eine definitorische Annäherung an den Begriff umso wichtiger. Außerdem soll hier ein Begriffsverständnis zugrunde gelegt werden, das auf wissenschaftlicher Erkenntnis und nicht auf der Interpretation von VLOPs und VLOSEs beruht. Denn deren Interpretation kann jederzeit einseitig verändert werden.

Grundlegend sind Desinformationen in der Forschung durch zwei Kriterien gekennzeichnet: Zum einen handelt es sich um Informationen, die faktisch nicht richtig oder irreführend sind. Und zum anderen werden diese falschen Informationen absichtlich verbreitet (Kessler, 2023). In Bezug auf den fehlenden Wahrheitsgehalt von Informationen kann darüber hinaus zwischen

verschiedenen sogenannten Informationsstörungen⁶ unterschieden werden (Dreyer et al., 2021; Kapantai et al., 2021). Im Gegensatz zu Desinformation liegt der Misinformation keine Intentionalität zugrunde, Schaden anzurichten. Zwar handelt es sich auch hier um falsche Informationen, diese werden jedoch nicht gezielt zur Destabilisierung verbreitet (Wardle & Derakhshan, 2017). Trotz dieser Begriffsvielfalt kann in der Anwendung auf den regulatorischen Kontext der EU eine gemeinsame definitorische Grundlage von Desinformation identifiziert werden, die durch die folgenden vier Charakteristika gekennzeichnet ist (Bleyer-Simon & Reviglio, 2024):

- es geht um nachweislich falsche oder irreführende Informationen
- diese haben das Potenzial, gesellschaftlichen Schaden anzurichten
- sie werden gezielt von bestimmten Akteur:innen gestreut
- die Verbreitung dient politischen oder ökonomischen Zwecken⁷.

Dennoch stellt diese Begriffsvielfalt für das Scoping Review eine Herausforderung dar, denn das Fehlen einer einheitlichen Definition in der herangezogenen Forschung erschwert die Zusammenführung der dabei gewonnenen Erkenntnisse. Hinzu kommt, dass im Zusammenhang mit Desinformationen auch häufig die Begriffe „Fake News“ und „Verschwörungserzählungen“ verwendet werden. Bei diesen handelt es sich um Subtypen von Desinformationen, die sich in ihrem Wahrheitsanspruch unterscheiden. Fake News meint die gezielte Verbreitung falscher Informationen mit einem wirtschaftlichen, politischen oder individuellen Ziel. Hingegen geht es bei Verschwörungserzählungen vorrangig darum, gezielt Narrative zu entwickeln, die ein exklusives Wissen über die „richtige“ Wahrheit zu einem bestimmten Thema suggerieren sollen (Butter, 2021). Deshalb wird hier ein weites Begriffsverständnis von Desinformationen zugrunde gelegt, das die o.g. vier Charakteristika beinhaltet, aber gleichzeitig keine Studien ausschließt, in denen nur einige dieser Charakteristika genutzt werden. Darüber hinaus nimmt das weite Begriffsverständnis in dieser Studie nicht nur Bezug auf einzelne Inhalte, die falsche oder irreführende Informationen beinhalten, sondern versteht diese im Kontext einer Vielzahl solcher Inhalte, die in größere Narrative der populistischen Online-Kommunikation eingebettet sind.

⁶ Zu diesen gehören, wie Kapantai et al. (2021) unterscheiden, verschiedene Typen von Desinformation: Bei „Clickbait“ handelt es sich beispielsweise um im Allgemeinen glaubwürdige Inhalte, die aber absichtlich übertriebene oder irreführende Überschriften und Thumbnails verwenden, um Nutzende auf die entsprechende Website zu locken. „Fabricated content“ hingegen sind Inhalte, die jeder faktischen Grundlage entbehren und deren Absicht es ist, zu täuschen und Schaden anzurichten. „Imposter content“ ist eine weitere Art von Desinformation, die beispielsweise Logos oder das Branding von echten, vertrauenswürdigen Quellen imitiert, um falsche oder irreführende Informationen zu verbreiten.

⁷ In Kapitel 4.2 werden die unterschiedlichen Motive genauer beleuchtet.

3.3 Die digitale Transformation politischer Öffentlichkeit

Der Erfolg des Internets und der digitalen Medien sowie die damit einhergehenden Veränderungen der Logik und der Struktur des Mediensystems verändern die diskursiven Machtstrukturen in der politischen Öffentlichkeit und somit in Politik und Gesellschaft.

War das Mediensystem vor dem Aufkommen des Internets durch die prä-digitale Logik der Massenmedien geprägt, interagieren und konkurrieren nun traditionelle Massenmedien und digitale Medien miteinander und haben sich zu einem neuen „hybriden Mediensystem“ entwickelt (Chadwick, 2013). Die traditionelle Massenmedienlogik wurde als „one-way“ und „top-down“ gerichtete Kommunikation charakterisiert, bei der die Informationsauswahl, -produktion und -verbreitung primär von Medienakteur:innen dominiert wird. Die traditionellen Medien fungieren hier als zentrale Informationsverteiler:innen an die Öffentlichkeit, was ihnen eine Position als „Gatekeeper“ von Informationsflüssen verliehen hat (Chadwick, 2013; Klinger & Svensson, 2015; Van Aelst & Walgrave, 2016). Im Kontext digitaler Medien hat sich dagegen eine sog. „network media logic“ und eine „many-to-many“ Kommunikationsumgebung etabliert, in der die benötigten Ressourcen für die Produktion und Verbreitung von Informationen dramatisch gesunken sind. Hierdurch erhalten im digitalen Raum neue Akteursgruppen die Möglichkeit, selbst medial-kommunikativ zu partizipieren. Informationen werden zunehmend in Echtzeit kommuniziert und konsumiert. Dabei fungieren soziale Plattformen und Suchmaschinen (sowie deren Plattform-Architektur und Algorithmen) als technische Vermittler und als Infrastruktur eines substantiellen Teils dieser Kommunikation (Klinger & Svensson, 2015; Thorson & Wells, 2016).

In diesem neuen hybriden Mediensystem sind viele unterschiedliche Kommunikator:innen aktiv, alte wie neue, darunter traditionelle Medien, „digital-born“ Nachrichtenportale, sog. „alternative“ und „hyperpartisan“ Nachrichtenkanäle, Journalist:innen, Influencer:innen, politische Parteien und Politiker:innen, Vertreter:innen von neuen und alten politischen Bewegungen, Aktivist:innen, zivilgesellschaftliche Gruppen und Akteur:innen, Unternehmen und viele andere (Jungherr et al., 2020). Traditionelle Medienakteur:innen bleiben in einer dominanten und mächtigen Position mit großen Reichweiten, werden jedoch stark von den neuen Kommunikator:innen beeinflusst und stehen mit ihnen im Wettbewerb (Jungherr et al., 2020; Behre et al (2024)). Eine zentrale Rolle spielt auch das Publikum und insbesondere aktiv partizipierende Bürger:innen, die sich online mit politischen Inhalten auseinandersetzen. Während diese Bürger:innen als einzelne Nutzer:innen selbst keine substantielle Reichweite auf den Plattformen aufbauen, beeinflussen ihre kumulierten Aktivitäten und Interaktionen (Kommentieren, Teilen und Liken) die Verbreitung von Information. Sie bieten zentrale Erfolgsmaße für die Reichweite der oben aufgeführten Kommunikator:innen an und liefern Feedback und Kommentare für diese (Bail, 2021; Jungherr et al., 2020). Infolge des Einflusses neuer digitaler Kommunikator:innen sowie dieser digital aktiven Bürger:innen und des Publikums im Allgemeinen entwickeln sich die früheren Privilegien der Massenmedien in Bezug auf Agenda-Setting und Framing zu einem netzwerkartigen Agenda-Setting und Framing, nicht

nur innerhalb der digitalen Medien, sondern im gesamten Mediensystem (Bennett & Pfetsch, 2018).



KURZ GEFASST: ALTERNATIVE UND HYPERPARTISANE MEDIENSEITEN

Alternative Medienseiten fungieren als Gegenpol zur hegemonialen Öffentlichkeit. Dazu gehören digitale Plattformen wie Social Media und Blogs, die der Förderung partizipativer Kommunikation dienen und alternative Perspektiven ermöglichen. Diese werden als Chance für deliberative Prozesse betrachtet, sind jedoch häufig negativ konnotiert, da sie als potenzielle Gefahr für die Demokratie wahrgenommen werden. Denn das Spektrum dieser Medien ist breit gefächert und reicht von kritisch-investigativen Formaten wie Krautreporter und Republik bis hin zu umstrittenen Publikationen wie Epoch Times und Unzensuriert. Ein zentrales Merkmal alternativer Medien ist die Bildung homogener Netzwerke, was zur Herausbildung alternativer Öffentlichkeiten beiträgt, in denen politische Botschaften gemeinsam konstruiert und verhandelt werden (Fenton & Barassi 2011, Schweiger 2017, Schwaiger 2022).

Hyperpartisane Medienseiten weichen von etablierten journalistischen Normen wie Objektivität und Unparteilichkeit ab, indem sie eine stark polarisierte Berichterstattung mit klarer ideologischer Ausrichtung betreiben. Sie bieten eine verzerrte Themenauswahl und verbreiten oft konfrontative Inhalte, die sich gegen die sogenannten „Mainstream-Medien“ richten. Charakteristisch für hyperpartisane Medien ist ihre explizite Parteilichkeit und die Abkehr von ausgewogener Berichterstattung. Sie nutzen Sensationalismus und Fehlinformationen, um Anti-Establishment-Narrative zu verstärken, wobei sie sowohl am linken als auch am rechten Rand des politischen Spektrums zu finden sind. Durch ihre Fokussierung auf ideologische Polarisierung tragen sie zur Fragmentierung der öffentlichen Debatte bei und untergraben das Vertrauen in traditionelle Institutionen (De Léon et al. 2024, Yan & Schroeder 2024).

Zwei zentrale Auswirkungen dieses digitalen Wandels auf die politische Öffentlichkeit stehen dabei besonders hervor. Erstens entsteht ein digitaler Möglichkeitsraum für politische Akteur:innen jenseits des tradierten politischen Spektrums, der es diesen ermöglicht, sich zu organisieren und den politischen Status quo herauszufordern (Bennett & Segerberg, 2012; Jungherr et al., 2019; Margetts et al., 2017). Durch die Nutzung digitaler Medien können diese „digitalen Neulinge“ Gemeinschaften aufbauen, Unterstützer:innen mobilisieren, neue Mitglieder rekrutieren und den öffentlichen Diskurs beeinflussen, ohne dabei auf etablierte mediale oder politische Strukturen angewiesen zu sein. Dieser Aufstieg digitaler Außenseiter:innen bringt eine neue Pluralität von verschiedenen Perspektiven in die Öffentlichkeit ein. Viele dieser politischen Newcomer fühlen sich demokratischen Werten

verpflichtet, aber das gilt nicht für alle. Tatsächlich scheinen insbesondere radikale und extreme Akteur:innen besonders erfolgreich darin zu sein, die geschwächte Gatekeeper-Position traditioneller Akteur:innen zu nutzen und den entstehenden neuen digitalen Möglichkeitsraum zu bespielen (Jungherr et al., 2020).

Zweitens sehen wir eine durch die digitale Transformation substantiell intensivierete „Aufmerksamkeitsökonomie“, mit erheblichen Implikationen für den politischen Diskurs. Das in den späten 1990er Jahren eingeführte Konzept (Franck 1998) umschreibt die Bedeutung von Aufmerksamkeit als Währung mit substanziellen Folgen für politische Diskurse. Die Masse an verfügbaren Informationen, in Kombination mit einer begrenzten menschlichen Aufmerksamkeit, hat den Wettbewerb um diese Aufmerksamkeit wesentlich verschärft (Webster, 2014). Dieser Wettbewerb wird von politischen Akteur:innen und von Medienakteur:innen angetrieben, die Aufmerksamkeit suchen, um ihre Botschaften zu kommunizieren, sowie von Plattformen und Medienakteur:innen, die auf Aufmerksamkeit für Werbeeinnahmen angewiesen sind.

Mit der erläuterten Verschiebung des Informationskonsums, weg von traditionellen Medienangeboten und hin zu digitalen Plattformen, geht eine wesentliche ökonomische Verschiebung einher: Für traditionelle Medien ist mit der Abwanderung von Werbegeldern zu digitalen Plattformen ein wesentlicher Teil ihrer bisherigen Finanzierungsbasis erodiert, was zu einem massiven Mediensterben in den letzten Jahrzehnten, insbesondere auch im lokalen und regionalen Raum beigetragen hat (Nielsen 2019). Die Plattformen können dagegen massive Einnahmen durch das Schalten von digitaler Werbung generieren und gehören damit teilweise zu den wertvollsten Unternehmen der Welt. Für Nutzende werden die Services der Plattform in der Regel kostenlos angeboten, „bezahlen“ tun sie jedoch oftmals mit ihren persönlichen Daten, die den Werbetreibenden zur Personalisierung der Werbung zur Verfügung gestellt werden (Makridis/Thayer 2023). Ein Teil der Einnahmen aus der geschalteten Werbung werden an Content-Creator:innen (bspw. Medienakteur:innen) weitergeben, wodurch für diese Monetarisierungsmöglichkeiten entstehen. Allerdings sind diese dadurch oft stark von den Plattformen abhängig und sind stark von potentiellen Änderungen der Community-Guidelines oder des Platform-Designs betroffen (siehe bspw. die Entscheidung von Meta, die Sichtbarkeit von News-Content zu reduzieren, wodurch die Nachrichtenseiten Einnahmeverluste zu verzeichnen haben).

Die Entwicklung digitaler Tracking-Systeme, die eine präzise Messung der Aufmerksamkeit ermöglichen, erlaubt es Inhaltsproduzent:innen und -verteiler:innen – einschließlich Politiker:innen, Medienhäusern und Plattformen –, ihre Inhalte zu optimieren, um sich an die Präferenzen von Nutzenden anzupassen und die Nutzendenaufmerksamkeit und das daraus resultierende Engagement zu maximieren (Jungherr et al., 2020). Häufig wird dabei auf kognitive Biases der Nutzenden zurückgegriffen. Studien in der Sozialpsychologie zeigen, dass Menschen dazu neigen, Informationen in ihrer Informationswahrnehmung und -verarbeitung zu bevorzugen, die ihre „In-Groups“ unterstützen, die emotional – insbesondere negativ

emotional – aufgeladen sind und moralisierte Informationen beinhalten (Brady et al., 2023)⁸. Da sich diese sogenannten Biases über Jahrtausende entwickelt haben, um menschliche Kooperation und kollektive Problemlösung zu fördern, und so eine wichtige Rolle für das evolutionäre Überleben gespielt haben, ist fraglich, ob sie das auch im hyperkompetitiven Umfeld der digitalen Aufmerksamkeitsökonomie weiterhin tun. Dies wird im Folgenden dargelegt.

3.4 Echokammern und Filterblasen

Eine viel diskutierte Thema ist die Entstehung möglicher Echokammern und Filterblasen im digitalen Raum (Hartmann et al., 2024; Bruns 2019). Eine grundlegende Herausforderung in dieser Diskussion ist die oft unscharfe Verwendung von Begriffen zur Beschreibung verwandter, aber doch unterscheidbarer Teilbereiche des vermuteten Phänomens. Rau und Stier (2019) differenzieren zwei unter dem Begriff der Echokammer-Hypothese zusammengefassten Teilhypothesen von 1) einer Fragmentierung der öffentlichen Aufmerksamkeit und 2) einer daraus resultierenden gesellschaftliche Polarisierung entlang politischer Vorprägungen selbst-selektierter Teilpublika.

Die erste Teilhypothese der Fragmentierung der öffentlichen Aufmerksamkeit beruht auf den Befürchtungen, dass im digitalen Raum Nutzende aufgrund psychologischer Dispositionen

1. Informationen primär entsprechend ihrer bereits existierenden politischen Präferenzen konsumieren, oder, vereinfacht ausgedrückt, politisch links eingestellte Menschen primär Informationen von politisch links verordneten Quellen beziehen, politisch rechts eingestellte Menschen von rechts verordneten Quellen,
2. sich online primär mit Gleichgesinnten umgeben, also nach dem gleichen Rechts-Links-Schema.
3. diese Vorlieben durch Empfehlungsalgorithmen nach dem Prinzip der Engagementmaximierung im Kontext der Aufmerksamkeitsökonomie weiter verstärkt werden (technische Ebene; siehe Filterblasendiskussion).

Empirisch konnte eine solche Fragmentierung nur teilweise nachgewiesen werden (Hartmann et al., 2024). Es trifft zu, dass politisch radikalere Nutzende häufig digital engagierter sind und dazu neigen, Inhalte zu teilen, die mit ihren eigenen Vorstellungen übereinstimmen (Aruguete et al., 2023). Dadurch können entsprechende digitale und politisch einseitige Communities entstehen. Diese Communities sind aber selten die einzige Informationsquelle im Repertoire der Nutzenden. Es ist also entscheidend, die Zusammensetzung der Informationsquellen von Nutzenden differenziert zu betrachten, da diese ein breites Spektrum unterschiedlicher Quellen

⁸ Es sollte hervorgehoben werden, dass, obwohl diese Biases gut dokumentiert sind, eine Vielzahl weiterer psychologischer Mechanismen und andere Faktoren die Informationsverarbeitung beeinflussen. Zum Beispiel Mechanismen wie epistemische oder offene Wachsamkeit, die die beschriebenen Verzerrungen möglicherweise (zumindest teilweise) ausgleichen könnten (Mercier, 2021; Sperber et al., 2010).

umfasst, von traditionellen bis stark parteiischen Medien. Dieses Muster bleibt auch in der sozialen Interaktion bestehen, es lassen sich oft politisch relativ homogene und teilweise auch radikale oder extreme Online-Communities finden, allerdings auch Orte, die von einer sehr diversen Nutzendenstruktur geprägt sind. Ob Empfehlungsalgorithmen bei der Frage geschlossener Gemeinschaften im Digitalen eine Rolle spielen, bleibt aufgrund der ungenügenden Datenlage (siehe Unterkapitel 5.3), der Diversität der Plattformen und der Algorithmen sowie der hohen Veränderungsrate über Zeit tendenziell unklar. Die vorhandenen Erkenntnisse deuten darauf hin, dass ein solcher selbstverstärkender Effekt durchaus vorkommen kann, seine Stärke und die Effekte lassen sich jedoch kaum bestimmen.

Hinsichtlich der Frage nach gesellschaftlicher Polarisierung als eine potenzielle Folge einer vermuteten Fragmentierung hat eine deutliche Ausdifferenzierung der Diskussion stattgefunden. So wird mittlerweile unter anderem zwischen (A) ideologischer, (B) affektiver, (C) themenbezogener, (D) diskursiver und (E) destruktiver Polarisierung im Kontext der digitalen Transformation politischer Öffentlichkeit gesprochen (für eine ausführliche Diskussion s. Kapitel 5.1). Mit Blick auf die Echokammer-Hypothese soll noch auf einen der wohl am stärksten polarisierenden Faktoren im digitalen Raum hingewiesen werden: das ständige Aufeinandertreffen antagonistischer politischer Akteur:innen, das durch die vernetzte Struktur des Internets erzeugt wird. Hierbei ist also nicht die Fragmentierung in getrennte Räume der wesentliche polarisierende Effekt, sondern das ständige Aufeinandertreffen in einem gemeinsamen Raum.



KURZ GEFASST: ECHOKAMMERN UND FILTERBLASEN

Die Diskussion über Echokammern und Filterblasen beinhaltet zwei Teiltheorien:

1. Nutzende konsumieren hauptsächlich Informationen, die ihren eigenen politischen Vorlieben entsprechen.
Allerdings zeigen Studien, dass Nutzende in der Praxis auf eine Vielfalt von Informationsquellen zugreifen.
2. Diese Fragmentierung fördert gesellschaftliche Polarisierung. Interessanterweise deuten Erkenntnisse darauf hin, dass das ständige Aufeinandertreffen unterschiedlicher politischer Akteure im selben digitalen Raum die gesellschaftliche Polarisierung verstärkt und nicht die Fragmentierung. Die Rolle von Empfehlungsalgorithmen bleibt dabei unklar.

Insgesamt bleibt die Frage nach den Auswirkungen solcher Phänomene auf die Gesellschaft ein komplexes und viel diskutiertes Thema.

4 REALISIERUNG SYSTEMISCHER RISIKEN AUF DIGITALEN DIENSTEN

Wie realisieren sich nun systemische Risiken, und hier vor allem Desinformationen im Kontext demokratischer Prozesse, auf digitalen Diensten? Dafür wird hier zunächst dargelegt, welche Akteur:innen an der Verbreitung von Desinformationen beteiligt sind und welche Motive diese haben. Anschließend werden prominente Verbreitungsstrategien und -mechanismen skizziert, bevor auf die Rolle von Suchmaschinen und Empfehlungssystemen sowie das Gefährdungspotenzial durch KI eingegangen wird. In einer abschließenden Reflektion werden die gegenwärtig größten Bedrohungen konsolidiert.

4.1 Hauptakteur:innen bei der Verbreitung von Desinformationen und ihre Motive

Da digitale Dienste in einem vielschichtigen Netzwerk von Institutionen und Akteur:innen verortet sind, erfolgt auch die Verbreitung von Desinformationen im Kontext dieser Wechselbeziehungen zwischen Plattformen, Politik, Medienhäusern, Organisationen der Infrastruktur und weiteren Akteur:innen. So sind vor allem sehr große Online-Plattformen (VLOPs) und -Suchmaschinen (VLOSEs) in sogenannte Plattform-Ökosystemen eingebettet, in denen beispielsweise Datenbanken und weitere Ressourcen geteilt werden: „The extent to which VLOPs and VLOSEs are interconnected by reliance on the same assets and services for their governance measures and prevention of harm could be a factor in the level of systemic risk” (Broughton Micova & Calef, 2023, S. 36). Dabei liegen – entsprechend der Begriffsannäherung an Desinformationen (siehe Kapitel 3.2) – bei der Verbreitung von Desinformationen verschiedene Motive zugrunde, die Desinformationskampagnen leiten. Diese sind vorrangig in politische (bzw. ideologische) und kommerzielle (bzw. ökonomische) Motive zu differenzieren. Dabei zu unterscheiden sind inner- und außerstaatlich ausgerichtete Motive (auch Foreign Information Manipulation and Interference, FIMI genannt).

Bei der politisch motivierten Verbreitung von Desinformationen ist das Ziel „unsere freie und demokratische Gesellschaft durch gezielte Fehlinformation zu verunsichern und zu verhindern, dass Menschen aufgeklärte Entscheidungen in ihrem politischen oder alltäglichen Handeln treffen können” (Kessler, 2023, S. 4). Dabei steht die Destabilisierung von Staaten oder politischen Systemen im Vordergrund, indem eigene politische Agenden legitimiert und

gleichzeitig die Agenden der Gegenseite delegitimiert werden. Dahinter steht die Annahme, dass die eigene Ideologie moralisch überlegen ist, und andere, von der eigenen Ideologie abweichende Perspektiven, durch die gezielte Manipulation von Narrativen als minderwertig dargestellt werden müssen (Hameleers et al., 2022). Politisch motivierte Verbreitung von Desinformationen wird sowohl von staatlichen als auch von politischen Akteur:innen eingesetzt, wobei sich die zugrundeliegenden Motive jeweils etwas unterscheiden. Allgemein nutzen staatliche Akteur:innen Desinformationen vorrangig, um nationale Interessen – oft im Bereich der nationalen Sicherheit oder der geopolitischen Macht – zu fördern und gleichzeitig die politische Einflussnahme anderer Staaten zu schwächen. Bei außerstaatlichen Akteur:innen erfolgt dies in Form von FIMI, einer “activity [which] is manipulative in character, conducted in an intentional and coordinated manner, by state or non-state actors, including their proxies inside and outside of their own territory” (European External Action Service, 2024). Ein prominentes Beispiel hierfür ist die sogenannte Doppelgänger-Kampagne, mit der seit 2022 aktiv pro-russische Narrative durch gefälschte Medienseiten, manipulierte Inhalte und automatisierte Accounts verbreitet werden mit dem Ziel, westliche Regierungen zu diskreditieren, zur Polarisierung der Gesellschaft beizutragen und die Unterstützung für die Ukraine zu schwächen (Bouchaud et al., 2024). Politische nationale Akteur:innen wie Parteien oder Kandidat:innen hingegen nutzen Desinformationen besonders häufig im Kontext von Wahlprozessen, um diese entsprechend ihrer eigenen Agenda durch Polarisierung und Manipulation zu beeinflussen (Mirza et al., 2023).

Die kommerziell bzw. ökonomisch motivierte Verbreitung von Desinformationen zielt auf die Profitmaximierung durch Aufmerksamkeit in der digitalen Ökonomie. Wie bereits beschrieben, hat die digitale Transformation der politischen Öffentlichkeit zu einer deutlich intensivierten Aufmerksamkeitsökonomie geführt. Die Möglichkeit, Inhalte mit digitaler Werbung zu verknüpfen und damit Geld zu generieren – abhängig vom Aufmerksamkeitsvolumen, den diese Inhalte erzeugen – ist ein wesentlicher ökonomischer Faktor im digitalen Mediensystem (Nielsen, 2019). Diese Mechanismen spielen auch bei der digitalen Verbreitung von Desinformationen eine wichtige Rolle: So gibt es Hinweise, dass einige Journalist:innen Fakten absichtlich falsch darstellen, um sogenannte „Infowars“ zu befeuern, was wiederum auf die Aufmerksamkeitsökonomie zielt. Dabei können auch finanzielle Motive eine Rolle spielen, indem Journalist:innen Desinformationen in ihre Berichterstattung einbauen, um Clickbait zu erzeugen oder ihre Karrierechancen zu verbessern (Lecheler & Egelhofer, 2022). Durch die Verbreitung von Desinformationen sollen so das Engagement mit bestimmten Inhalten und damit auch die Werbeeinnahmen erhöht werden. Als Beispiel hierfür wird in der Literatur Facebook genannt, wo Mikrotargeting und algorithmische Verstärkung die Verbreitung von Desinformationen fördern soll (Hameleers et al., 2022). Ähnliches sieht die Forschung auch auf der Plattform Telegram, wo verschiedene Akteur:innen aus dem extremistischen und/oder verschwörungsideologischen Spektrum die von ihren Inhalten generierte Aufmerksamkeit zur Monetarisierung nutzen (Marcks, 2023). Hierbei werden auch die in den Vorarbeiten dargelegten kognitiven Mechanismen menschlicher Informationswahrnehmung und -

verarbeitung genutzt, um mit negativ-emotionalisierten und gruppenidentitätsbasierten Inhalten Aufmerksamkeit zu generieren (Jost & Dogruel, 2023).

Der Blick in die Berichte zur Europawahl 2024 von Meta und TikTok (Meta, 2024; TikTok, 2024) zeigt, dass zwischen inländischen und ausländischen Akteur:innen unterschieden werden muss. Bei Meta zeigt sich beispielsweise Russland als staatlicher Akteur als einer der zentralen ausländischen Akteur:innen bei der Verbreitung von Desinformationen. In Bezug auf die Doppelgänger-Kampagne identifizierte Meta die Operation bereits 2022 und ordnete sie russischen Unternehmen zu, die EU-Sanktionen unterliegen. Trotz Maßnahmen ist die Operation weiterhin aktiv. Das Risiko wird von Meta als gering eingeschätzt, obwohl die Aktivitäten weitreichend und intensiv sind (Bouchaud et al., 2024). Bei den inländischen Akteur:innen spielen insbesondere Politiker:innen und politische Parteien, Aktivist:innen und Protestgruppen sowie Blogger:innen und Influencer:innen eine wichtige Rolle bei der Verbreitung von Desinformationen. In den Berichten von Meta zeigte sich zudem, dass die meisten EU-bezogenen koordinierten inauthentischen Aktivitäten vor den EU-Wahlen 2024 inländischer Natur waren.

4.2 Technische Verbreitungsmöglichkeiten von Desinformationen in Online-Suchmaschinen und auf Online-Plattformen

Die genannten Hauptakteur:innen nutzen zur Verbreitung von Desinformationen auf Online-Plattformen und Online-Suchmaschinen verschiedene Verbreitungsstrategien („dissemination strategies“), die auf einer Verbindung unterschiedlicher technischer Möglichkeiten und menschlicher Verhaltensweisen basieren (Barman et al., 2024). Als Haupt-Verbreitungsmöglichkeit für Desinformationen ist hier vor allem strategische Manipulation, etwa durch staatliche Akteur:innen, Botnetzwerke und Troll-Subkulturen, zu nennen. Im Folgenden werden die unterschiedlichen Verbreitungsstrategien in Online-Suchmaschinen und auf Online-Plattformen nachgezeichnet.

4.2.1 Verbreitungsstrategien in Online-Suchmaschinen

Suchmaschinen werden als Quelle aktueller Nachrichten und Informationen ein insgesamt größeres Vertrauen entgegengebracht als Social-Media-Inhalten. Deshalb kommt Suchmaschinen eine besondere Rolle als Intermediär in gesellschaftlichen Debatten zu (Fletcher & Nielsen, 2018; Uрман & Makhortykh, 2023). Die Verwendung von Suchmaschinen zur Überprüfung von Nachrichten führt mitunter zu einem verstärkten Glauben an Fehlinformationen (Aslett et al., 2024). Dies lässt sich darauf zurückführen, dass Suchergebnisse, insbesondere bei aktuellen Nachrichten, mitunter irreführende Inhalte aus Quellen geringer Qualität priorisieren, die Fehlinformationen verstärken (Aslett et al., 2024). Dies lässt sich darauf zurückführen, dass Suchmaschinen bei solchen Suchbegriffen, zu denen keine seriösen Informationen vorliegen, mitunter darauf angewiesen sind, Informationen aus zweifelhaften Quellen anzuzeigen.

Desinformationen in Suchmaschinen sind stark sichtbar, weil die Verbreiter:innen von Desinformationen gezielt auf Techniken der Suchmaschinenoptimierung (SEO) zurückgreifen. Sie verwenden bestimmte Phrasen und Schlüsselwörter, um Suchergebnisse zu manipulieren und so Desinformation in den Suchergebnissen höher zu platzieren (Williams & Carley, 2023). Das Problem wird dadurch verstärkt, dass sich viele Nutzende, insbesondere solche mit geringerer digitaler Kompetenz, auf simple Suchbegriffe wie Überschriften oder URLs verlassen. Diese Suchanfragen liefern häufig unzuverlässige Ergebnisse, da sie Ausdrücke enthalten, die ausschließlich von Quellen geringer Qualität verwendet werden (Slechten et al., 2021). Wird zum Beispiel über die sozialen Medien der Verschwörungsmythos verbreitet, dass die Großväter führender deutscher Politiker eine Nazivergangenheit haben, können Suchmaschinen unseriöse Webseiten aufgreifen, welche diese Behauptung enthalten und ausführen, solange keine seriösen Quellen ihr explizit widersprechen.

Eine weitere Herausforderung ist die intransparente Sortierung der Quellenrangfolge innerhalb derselben Suchmaschine (Haas & Unkel, 2017). Dies kann dazu führen, dass Nutzenden unterschiedliche Informationen angezeigt werden und die Auswirkungen irreführender Quellen, die in einigen Fällen priorisiert werden könnten, weiter verstärken (Makhortykh & Wijermars, 2021). Zusammenfassend kann festgehalten werden, dass Suchmaschinen zur Verbreitung von Fehlinformationen beitragen können, obwohl sie als Werkzeuge zur Informationsbeschaffung und -überprüfung dienen sollten. Dies geschieht, indem sie Nutzende zu Quellen von geringer Qualität leiten und bestehende Vorurteile verstärken.

4.2.2 Verbreitungsstrategien auf Online-Plattformen

Desinformationen stellen insbesondere dann ein systemisches Risiko dar, wenn sie viele Nutzende erreichen (siehe z.B. Kessler, 2023). Dann nämlich kann das koordinierte und häufig als inauthentisch eingestufte Verhalten von Social-Media-Accounts zur Verbreitung von desinformierenden Inhalten durch das organisierte und nahezu zeitgleiche Posten und Liken von Inhalten der Eindruck erwecken, dass diese Inhalte besonders populär sind und eine vermeintlich breite Unterstützung erhalten (z.B. Aimeur et al., 2023; Cinelli et al., 2022; de-Lima-Santos & Ceron, 2023; Righetti et al., 2022).

Diese sind auch in den Berichten von Meta und TikTok zur Europawahl 2024 zu finden (Meta, 2024; TikTok, 2024). Bei Meta zeigte sich beispielsweise, dass die meisten der EU-bezogenen CIBs (Coordinated Inauthentic Behaviour), die im Vorfeld der EU-Wahlen 2024 unterbrochen wurden, inländischer Natur waren. Die Bedrohungen durch ausländische Akteur:innen – darunter die russische Doppelgänger-Kampagne⁹ – und eine Handvoll anderer Gruppen, konzentrierten sich in erster Linie darauf, die Unterstützung für die Ukraine in den EU-Mitgliedstaaten zu untergraben, und nicht direkt auf die europäischen Parlamentswahlen. Diese Gruppen setzten oft Taktiken ein, die eine nicht authentische Verstärkung von authentischen

⁹ Das sogenannte Smash-and-Grab-Projekt, welches sich vor allem auf das Spoofing von Webseiten konzentriert. https://cemas.io/publikationen/fortsetzung-folgt-doppelgaenger/241114_cemas_report_06_de_digital.pdf

Konten oder Seiten einheimischer Politiker:innen durch Likes, Shares und Kommentare beinhalteten. So erschienen die Konten oder Seiten populärer, als sie in Wirklichkeit waren. Es lassen sich jedoch keine eindeutigen Anzeichen dafür finden, dass diese Gruppen beim authentischen Publikum überdurchschnittlich viel Anziehungskraft erlangen (Geers et al., 2024; Meta, 2024).

Plattformübergreifend zeigt sich in diesen Berichten, dass vor allem das Erstellen fiktiver Nachrichtenwebsites, die Nachahmung seriöser Medien („Doppelgänger“), die künstliche Verstärkung authentischer Inhalte sowie der zunehmende Einsatz von KI-generierten Inhalten häufig genutzte Taktiken zur Verbreitung von Desinformationen sind. Dabei kann koordiniertes inauthentisches Online-Verhalten auf organisierte Gruppen zurückgeführt werden, die mittels strategischer Absprachen und der Nutzung mehrerer Social-Media-Accounts pro Gruppenmitglied Inhalte massenhaft verbreiten. Bei diesen Botnetzwerken handelt es sich um Softwaresysteme, die die Online-Aktivitäten zahlreicher gefälschter Konten steuern, um einerseits irreführende Inhalte zu verbreiten und andererseits bestimmte Inhalte zu teilen und zu bewerben (Barman et al., 2024). Die korrekte Bestimmung solcher Botnetzwerke ist allerdings methodisch herausfordernd, da bestehende Ansätze noch nicht ausgereift sind und teilweise Social-Media-Accounts fehlerhaft einstufen (de-Lima-Santos & Ceron, 2023)¹⁰. Gerade besonders fortgeschrittene technische Verfahren wie generative KI erschweren die zweifelsfreie Detektion von Bots. Entsprechend ist die korrekte Zuschreibung der Urheberschaft eines identifizierten inauthentischen koordinierten Verhaltens ebenso schwierig wie die Einschätzung, ob es sich bei diesem beobachteten Verhalten um eine authentische Graswurzelbewegung oder einen strategischen Manipulationsversuch durch organisierte politische Gruppen handelt (Cinelli et al., 2022; Manocci et al., 2024).

Ein weiterer Verbreitungsweg von Desinformationen, der insbesondere in Online-Subkulturen Anwendung findet, ist das sogenannte Trollen: die Verbreitung provokativer oder irreführender Inhalte zur Erzeugung von Aufmerksamkeit (Van Raemdonck & Meyer, 2024). So entstehen Verschwörungserzählungen häufig auf Plattformen, die für ironische und oft toxische Kommunikation bekannt sind. Diese Theorien können durch einen Prozess der „Normifizierung“ auch auf Nutzende außerhalb der ursprünglichen Subkultur übergehen, die die Inhalte dann ernsthaft weiterverbreiten und sich ideologisch damit identifizieren (Van Raemdonck & Meyer, 2024).

4.3 Die Rolle von Empfehlungssystemen auf Online-Plattformen und in Online-Suchmaschinen bei der Verbreitung von Desinformationen

Empfehlungssysteme sind darauf ausgelegt, das Engagement der Nutzenden zu steigern. Gleichzeitig steigt dadurch – zumindest lokal – das Potenzial, geschlossene Informationsökosysteme zu schaffen. Dieser Effekt ist besonders auf Social-Media-Plattformen

¹⁰ Zwar gibt es Erkenntnisse, wie sich automatisierte Botnetzwerke im Nutzungsverhalten von Menschen betriebenen Accounts unterscheiden, allerdings lassen sich daraus noch keine absolut fehlerfreien Identifizierungen von Botnetzwerken ableiten.

deutlich. Bei diesen kann es eine kontinuierliche Auswertung von Interaktionsmustern in Kombination mit dem strategischen Handeln staatlicher Akteur:innen geben. Dies kann dazu führen, dass algorithmische Empfehlungssysteme gezielte Desinformationskampagnen begünstigen (Guess et al., 2023). In verschiedenen Studien konnten mehrere miteinander verbundene technische Mechanismen identifiziert werden, die zur algorithmischen Verbreitung falscher Informationen über digitale Plattformen beitragen. Untersuchungen zu YouTube haben beispielsweise nachgewiesen, dass der Algorithmus der Plattform in Bezug auf verschwörungsbezogene Inhalte systematisch jenen Nutzenden Videos empfahl, die sich zuvor mit ähnlichem Material beschäftigt hatten. Dieses Muster legt nahe, dass Empfehlungssysteme - gerade weil es ihr beabsichtigter Zweck ist, Nutzenden für sie individuell relevante Inhalte zu zeigen - auch die Verbreitung von Fehlinformationen fördern können, indem sie die Sichtbarkeit von sachlichen Informationen oder ausgleichenden Perspektiven einschränken.

Dazu ist in der öffentlichen Debatte über die Phänomene der Filterblase und Echokammern (González-Bailón & Lelkes, 2023; Steppat et al., 2022) entstanden. Die Existenz und der Stellenwert dieser sogenannten Filterblasen sind in der Forschung umstritten, wie oben dargestellt (Ross Arguedas et al., 2022). Wissenschaftliche Erkenntnisse zeigen jedoch tiefgreifende Asymmetrien bei der Interaktion und Zustimmung zu Social-Media-Inhalten, je nach sozialer Identität oder der Parteizugehörigkeit der Nutzenden (González-Bailón & Lelkes, 2023). So besteht das Risiko, dass die Verwendung historischer Interaktionsdaten und Informationen zur Soziodemografie der Nutzenden, die in die algorithmische Personalisierung einfließen, homogene Informationsumgebungen schaffen (Guess et al., 2023). Denn Nutzende haben in der Regel nur eingeschränkte Kontrolle über ihre Informationsverfügbarkeit, da Algorithmen Inhalte automatisch auf der Grundlage abgeleiteter Präferenzen filtern und organisieren. Diese Einschränkung kann die Fähigkeit der Nutzenden beeinträchtigen, aktiv nach unterschiedlichen Perspektiven zu suchen und bestehende Überzeugungen in Frage zu stellen. Das verstärkt möglicherweise die Effekte der homogenen Informationsverfügbarkeit (Figà Talamanca & Arfini, 2022).

Hinzu kommt, dass das koordinierte, inauthentische und massenhafte Posten sowie Liken durch Social-Media-Accounts durch Empfehlungsalgorithmen verstärkt wird. Denn diese priorisieren vorrangig polarisierende oder virale Inhalte, unabhängig von deren Richtigkeit (Van Raemdonck & Meyer, 2024). Dabei sortieren die Empfehlungsalgorithmen die riesigen Mengen an Inhalten und präsentieren sie den Nutzenden der Plattformen in einer überschaubaren und ansprechenden Form. Dies erfolgt auf Basis gesammelter Daten über die Nutzenden, wie ihre Zugehörigkeit zu einer bestimmten Bevölkerungsgruppe, ihr früheres Verhalten auf der Plattform und/oder das Verhalten ihrer Familienmitglieder oder ihres Freundeskreises (Sun, 2023).

Ein weiterer wichtiger technischer Faktor sind die sogenannten „Datenlücken“ (data voids). Diese bezeichnen Leerstellen im Inhaltsangebot großer Plattformen. Datenlücken treten auf, wenn maßgebliche Quellen zu bestimmten Themen noch keine substantziellen Inhalte produziert haben. Dadurch entsteht die Möglichkeit, dass diese Lücken durch Fehlinformationen

vermeintlich „geschlossen“ werden. Dieses Problem wird besonders akut bei neuen oder umstrittenen Themen, bei denen glaubwürdige Quellen möglicherweise keine umfassenden Analysen veröffentlichen. Nutzende, die sich auf einfache Suchbegriffe wie Überschriften oder direkte URLs verlassen, sind stärker gefährdet, auf Datenlücken zu stoßen, da diese Suchmuster häufig mit der von weniger zuverlässigen Quellen bevorzugten Terminologie übereinstimmen.

Die Intransparenz und eingeschränkte Nutzerkontrolle algorithmischer Systeme stellen zusätzliche technische Herausforderungen dar. Denn die Empfehlungsalgorithmen großer Plattformen wie Google, YouTube und Facebook sind in ihrer Funktionsweise intransparent. Das macht es schwierig, ihre Entscheidungsprozesse und potenziellen Verzerrungen zu verstehen. Dieser Mangel an Transparenz schränkt die Fähigkeit der Nutzenden ein, die Zuverlässigkeit empfohlener Informationen zu bewerten und algorithmische Verzerrungen zu identifizieren.

4.4 Bewertung des Gefährdungspotentials durch KI und durch diese generierten Inhalte

KI-Systeme steuern die Sichtbarkeit und Verbreitung von Informationen, Fehlinformationen und Desinformation auf Plattformen. Kein VLOP/VLOSE kommt ohne KI-unterstützte Empfehlungsalgorithmen in den Content-Governance-Systemen aus. Gleichzeitig können große Sprachmodelle (generative KI) zur Produktion von Desinformation eingesetzt werden. Generative KI bezeichnet dabei Technologien, die mit Hilfe maschinellen Lernens neue Inhalte wie Texte, Bilder, Audio oder Videos generieren, indem sie innerhalb großer Datenmengen wiederkehrende Muster identifizieren und anschließend auf deren Grundlage eigenständig Output produzieren. Zu den am weitesten verbreiteten Anwendungen gehören chat-basierte Systeme, die es Nutzenden ermöglichen, durch gezielte Eingaben („Prompts“) auf einfache Weise Informationen zu suchen, Bilder oder Memes zu erstellen oder multimediale Inhalte wie Videos zu generieren, ohne dass technische Vorkenntnisse erforderlich sind. Diese Systeme unterstützen verschiedene Sprachen und sind zunehmend in der Lage, komplexe Aufgaben – wie mehrstufige Desinformationskampagnen – auszuführen. Wenn Plattformen nicht ausreichend dafür Sorge tragen, dass künstlich generierte Inhalte als solche erkennbar sind oder die Kommunikationsmuster und -modi von KI-Akteur:innen erkannt werden können, kann generative KI deshalb negative Effekte auf die Informationsqualität haben. Gleichzeitig können KI-Anwendungen auch als Mittel im Kampf gegen Desinformationsinhalte eingesetzt werden.

Erste repräsentative Studien zur Nutzung von chat-basierten KI-Anwendungen wie ChatGPT zeigen, dass diese Tools von Bürger:innen in Deutschland mehrheitlich noch nicht als Quelle für politische Informationen genutzt werden. Wuttig und Kessler (2024) identifizieren hierbei ein klares Altersgefälle: In den Generationen Z (etwa Geburtsjahrgänge 1997-2012) und Y (etwa 1980 bis 1997) haben 36 % bzw. 28 % schon einmal chat-basierte KI-Anwendungen wie ChatGPT als Quelle für politische Informationen genutzt – in der Generation X (etwa 1965-1980) und bei den Babyboomern (etwa 1946-1964) lediglich 15 % bzw. 12 % (Wuttig & Kessler, 2024). Dennoch zeigen sich alle Generationen skeptisch hinsichtlich deren Informationsgehalts. Sie sind sich darüber hinaus mehrheitlich der Gefahr des missbräuchlichen Gebrauchs von KI-Technologien zur Manipulation und Fälschung von Inhalten (73 %) bewusst. Auch über eine

mögliche Verstärkung von Ungleichheiten und Diskriminierung durch fehlerhafte KI-Programmierung (66 %) sind sie sich im Klaren (Wuttig & Kessler, 2024). Auch Reiss und Kolleginnen (2025) stellen fest, dass chat-basierte KI-Anwendungen als Nachrichtenquelle in Deutschland derzeit nur eine geringe Relevanz haben. Die Mehrheit der Befragten einer repräsentativen Stichprobe von Online-Nutzenden ab 16 Jahren gab an, diese Tools entweder nie oder nur sehr selten zur Information über politische Themen zu verwenden. Allerdings zeigt sich, dass insbesondere Personen in schulischer oder betrieblicher Ausbildung signifikant häufiger auf diese Anwendungen zur politischen Informationsgewinnung zurückgreifen. Zudem nutzt diese Gruppe chat-basierte KI-Plattformen generell häufiger und primär zu Weiterbildungszwecken. Obwohl die Nutzung für politische Themen in der Gesamtbevölkerung aktuell auf einem niedrigen Niveau liegt, lässt sich vermuten, dass sich das ändern wird: Vor allem Personen, die durch den Einsatz chat-basierter KI im Rahmen ihrer Weiterbildung bzw. beruflichen Nutzung positive Erfahrungen gemacht haben, könnten künftig ihre Nutzung auch auf weitere Anwendungsbereiche wie die politische Informationssuche ausweiten.

Daneben können Nutzende – wissentlich oder unwissentlich – auch über journalistische Medienangebote in Kontakt mit KI-generierten Inhalten kommen. Bereits jetzt setzen diverse Medienhäuser in Deutschland und Europa entsprechende Technologien ein (Beckett & Yaseen, 2023; KI-Reifegrad Report, 2024). Der AI Act der Europäischen Union (Verordnung über künstliche Intelligenz, 2024) sieht klare Anforderungen und Verpflichtungen für Entwickler:innen und Betreiber:innen von KI-Systemen vor, der eine Kennzeichnungspflicht für den Einsatz automatisierter Systeme vorschreibt.

In der Produktion wie in der Moderation und Kontrolle potenziell desinformierender Inhalte können KI-Anwendungen eingesetzt werden. So nutzen Online-Plattformen schon jetzt KI, um Richtlinien zur Inhaltsmoderation umzusetzen (Dergacheva et al., 2023). Inhaltsmoderation umfasst dabei das Erkennen und Entfernen oder Markieren unangemessener, schädlicher oder illegaler Inhalte basierend auf vordefinierten Kriterien. Das Gutachten der Goldmedia GmbH im Auftrag der Bundesnetzagentur (Bundesnetzagentur 2023) analysiert den aktuellen Stand der Inhaltsmoderation bei Hostingdiensten im Kontext der EU-Verordnung zur Bekämpfung terroristischer Online-Inhalte (TCO-VO). In dem Gutachten wird betont, dass automatisierte Moderationsverfahren, wie Wortfilter, Hash-Vergleiche für Bild- und Toninhalte sowie KI-basierte Sentimentanalysen, für Textinhalte vorrangig zur Unterstützung und Vorbereitung von Löschentscheidungen genutzt werden. Die teilweise manuelle Moderation bleibt unerlässlich, insbesondere um den kulturellen Kontext von Äußerungen zu verstehen. Die Studie hebt auch hervor, dass automatisierte Lösungen insbesondere bei der Textanalyse überzeugend arbeiten, während ihre Anwendung bei Video- oder Audioinhalten sowie bei Live-Streams noch begrenzt ist. In personalisierten Informationsumgebungen wie Suchmaschinen und sozialen Medien nutzen Plattformen KI-basierte Systeme, um Inhalte zu kuratieren, zu selektieren und zu organisieren. Diese Systeme sorgen dafür, dass nur Inhalte verbreitet werden, die zuvor eine Moderationsprüfung durchlaufen haben. Dabei bestimmen sie, welche Inhalte welchen Nutzenden angezeigt werden, indem sie personalisierte Empfehlungen auf Basis individueller

Präferenzen generieren. Diese Präferenzen werden aus dem bisherigen Verhalten der Nutzenden abgeleitet (Forum on Information and Democracy, 2023).

Die potenziellen Auswirkungen generativer KI auf Desinformationen lassen sich in vier Kategorien einteilen: höhere Quantität, bessere Qualität, verstärkte Personalisierung und „Halluzination“, also die Generierung plausibler, aber falscher Informationen (Xu et al., 2024). Die Messung des Umfangs von KI-generierter und -verbreiteter Fehlinformation sowie der Auswirkungen von Desinformationskampagnen gestaltet sich jedoch schwierig. Aufgrund des fehlenden Datenzugangs ist schwierig, Daten zu identifizieren, zu sammeln und zu analysieren, die die alltäglichen Online-Erfahrungen der Menschen in personalisierten Informationsumgebungen vollständig widerspiegeln (Haidt & Bail, 2024). Die vorhandenen Befunde deuten jedoch darauf hin, dass der Umfang von KI-generierter und -verbreiteter Fehlinformation in den fünf Jahren bis 2023 erheblich zugenommen hat, auch wenn unklar ist, welchen Einfluss KI-generierte Desinformation konkret auf Meinungsbildungsprozesse hat (Guess et al., 2023).

Deepfakes stellen einen Sonderfall der KI-generierten Desinformation dar. Dabei handelt es sich um realistisch wirkende audiovisuelle Inhalte, die mit Hilfe von KI erzeugt werden. Birrer und Just (2024) zeigen in ihrer Metaanalyse, dass etwa 40 % der Teilnehmenden in empirischen Studien Deepfakes nicht als manipuliertes Material identifizieren konnten und ihre Fähigkeiten zur korrekten Identifikation überschätzen. Eine jüngst von Bitton und Kolleg:innen (2024) in Deutschland durchgeführte Studie kam zu dem Ergebnis, dass jüngere Altersgruppen, höher gebildete und männliche Personen sowie Menschen mit digitaler Kompetenz, die häufig soziale Medien nutzen, über umfangreicheres Wissen zur Deepfake-Technologie verfügen. Alters- und geschlechtsspezifische Unterschiede hinsichtlich der Fähigkeit, Deepfakes zu erkennen, werden dabei auf einen häufigeren Kontakt mit Deepfakes und einen insgesamt erfahreneren Umgang mit digitalen Medien zurückgeführt und diskutiert (Bitton et al., 2024; Hashmi et al., 2024; Sippy et al., 2024).

Die insgesamt geringe Erkennungsrate von Deepfakes lässt sich durch das grundsätzliche Vertrauen der Menschen in audiovisuelle Inhalte erklären. Dieses Vertrauen lässt zunächst nicht vermuten, dass die Inhalte überzeugend manipuliert oder vollständig künstlich generiert sein könnten (Hameleers & Marquart, 2023). Die fortschreitende Verbesserung der Deepfake-Qualität stellt in diesem Zusammenhang ein erhebliches Problem dar: Artefakte wie inauthentische Mimik, unnatürliche Sprachrhythmen oder eine geringe Bildqualität, die Deepfakes bisher (leichter) erkennbar machten (Birrer & Just, 2024; Jin et al., 2023), werden durch den technischen Fortschritt zunehmend beseitigt (Patel et al., 2023).

Insbesondere im Kontext demokratischer Prozesse wird die Wirkung politischer Deepfakes, wie beispielsweise manipulierte Reden von Politiker:innen, zunehmend als Bedrohung für die Demokratie diskutiert. Denn politische Deepfakes können genutzt werden, um Desinformation zu verbreiten und das Vertrauen in öffentliche Institutionen zu untergraben oder Wahlen zu manipulieren (Mubarak et al., 2023). Neben der audiovisuellen Qualität erhöht auch die

inhaltliche Plausibilität die Glaubwürdigkeit von Deepfakes: Eine Studie zeigte, dass hyperrealistische Deepfakes mit glaubwürdigen Inhalten überzeugender wirken als inhaltlich unrealistische Manipulationen (Hameleers et al. 2024a). Daneben scheint auch die Resonanz in Online-Kontexten, die Followeranzahl der Quelle, die Popularität eines Deepfakes sowie eine hohe Bildqualität die Glaubwürdigkeit und Überzeugungskraft von Deepfakes positiv zu beeinflussen (Hameleers et al., 2024b; Jin et al., 2023). Ob Deepfakes insgesamt überzeugender sind als textliche Desinformation ist jedoch umstritten: Einzelne Studien deuten darauf hin, dass Deepfakes zu einer Diskreditierung der dargestellten Politiker:innen führen können oder im Vergleich mit anderen als lebendiger und glaubwürdiger wahrgenommen werden. Andere Untersuchungen kommen dagegen zu dem Ergebnis, dass Deepfakes keine größere Wirkung entfalten (Birrer & Just, 2024; Lee & Shin, 2022).

Obwohl die Befürchtung einer massenhaften Manipulation der Bevölkerung durch Deepfakes aktuell nicht empirisch bestätigt werden kann (Birrer & Just, 2024; Hameleers et al., 2024a, 2024b), bleibt abzuwarten, wie sich die Weiterentwicklung von immer realistischeren Deepfakes auf deren Überzeugungskraft auswirkt und ob die Fähigkeit der Menschen, diese zu identifizieren, in gleichem Maße zunimmt.

Neben der umstrittenen direkten Wirkung von Deepfakes auf (politische) Meinungen und Einstellungen werden sie als potenzielle Gefahr für das gesellschaftliche Klima und das allgemeine Vertrauen in Medien gesehen (Birrer & Just, 2024). Der bewusste Kontakt mit Deepfakes, etwa durch Labels zur Kennzeichnung von Desinformation, kann die Unsicherheit der Rezipierenden gegenüber Medieninformationen verstärken und das allgemeine Medienvertrauen mindern (Bitton et al., 2024; Lewis et al., 2023; Weikmann et al., 2024). Hameleers und Marquart (2023) haben gezeigt, dass inkorrekte Kennzeichnungen authentischer Inhalte als Desinformation das Vertrauen in den Nutzen einer Kennzeichnung nachhaltig untergraben. Sie warnen, dass böswillige Akteur:innen unter dem Vorwand der Faktenprüfung wahre Inhalte gezielt delegitimieren könnten.



KURZ GEFASST: SORGEN UM ERODIERENDES VERTRAUEN

KI-Systeme beeinflussen die Verbreitung von Informationen und Desinformation auf digitalen Plattformen und werden sowohl zur Moderation als auch zur Erstellung manipulativer Inhalte eingesetzt. Besonders generative KI, darunter Deepfakes, kann das Vertrauen in Medien gefährden. Während der direkte Einfluss noch nicht abschließend erforscht ist, zeigt sich eine wachsende Besorgnis über die Auswirkungen auf Desinformation und öffentliche Meinungsbildung.

5 AUSWIRKUNGEN SYSTEMISCHER RISIKEN AUF DIE GESELLSCHAFT

Nachdem im vorherigen Kapitel die Akteur:innen, Verbreitungsmöglichkeiten und Gefährdungspotentiale zur Verbreitung von Desinformationen auf Online-Plattformen und Online-Suchmaschinen herausgearbeitet wurden, werden im Folgenden die Auswirkungen von Desinformationen im Kontext demokratischer Prozesse auf die Gesellschaft beleuchtet. Dazu werden zunächst potenzielle soziale und politische Auswirkungen aufgezeigt, bevor diskutiert wird, inwiefern eine (Mit-)Ursächlichkeit zwischen der Verbreitung von Desinformationen und den genannten Auswirkungen nachgewiesen werden kann. Darauf aufbauend werden verschiedene Möglichkeiten der Feststellung und Quantifizierung der Auswirkungen dargelegt. In Zusammenhang mit der Frage, wie Nutzende die Gefahren von Desinformationen bewerten, werden Erkenntnisse über neueste Entwicklungen konsolidiert und Möglichkeiten zur Begrenzung dieser Auswirkungen aufgezeigt.

5.1 Soziale und politischen Auswirkungen von Desinformationskampagnen und anderen Methoden der Beeinflussung

Eine der größten Herausforderungen der digitalen Transformation politischer Öffentlichkeit für Politik und Gesellschaft ist die konfliktorientierte und negativ-emotionalisierte Verschiebung des öffentlichen Diskurses. Diese resultiert aus den in Kapiteln 3 und 4 beschriebenen menschlichen Biases in der Informationswahrnehmung und -verarbeitung und den Regeln und automatisierten Content-Governance-Systemen im Dienst der digitalen Aufmerksamkeitsökonomie. Wie schon dargelegt, zeigen Studien aus der Sozialpsychologie, dass Menschen dazu neigen, tendenziell Kommunikationsinhalte wahrzunehmen und zu verarbeiten, die ihre Gruppenidentität ansprechen, die negativ-emotional aufgeladen sind und moralische Botschaften enthalten. Zusätzlich zu diesen Biases fördert auch die digital verschärfte Aufmerksamkeitsökonomie einen diskursiven Stil im digitalen Raum. Dieser ist durch Gruppenzugehörigkeiten, Konflikte zwischen diesen Gruppen, emotionalisierten Auseinandersetzungen und moralisierender Sprache geprägt (Brady et al., 2023).

Anstelle sachbezogener, lösungsorientierter Debatten fokussiert sich dieser politische Diskurs häufig auf negativ-emotional aufgeladene Unterscheidungen zwischen „wir“ und „die“, was die Gruppenidentitäten stärkt (Brady et al., 2019). Die Vielfalt und Unterschiedlichkeit der Stimmen im digitalen Raum, insbesondere auch die hohe Sichtbarkeit radikaler und extremer Akteur:innen sowie die vernetzte Struktur des digitalen Raums verstärken die genannten Tendenzen zusätzlich. Entgegen der weit verbreiteten Echokammer-Hypothese (siehe Kapitel

3.4) zeigen Untersuchungen, dass Polarisierung in digitalen Umgebungen häufig durch kontinuierliche, unmoderierte Konfrontation mit gegensätzlichen politischen Perspektiven zunimmt (Bail et al., 2018; Bruns, 2019; Rau & Stier, 2019). Sogenannte Polarisierungsunternehmer:innen verstärken diese Antagonismen bewusst, da konfliktgeladene und negativ-emotionale Kommunikation eine erfolgreiche Agenda-Setting-Strategie in der digitalen Aufmerksamkeitsökonomie ist und dazu beiträgt, ein günstiges politisches Klima für ihre Agenda zu schaffen (Jungherr et al., 2020; Mau et al., 2024).

Diese konfliktorientierte und negativ-emotionalisierte Verschiebung des öffentlichen Diskurses durch digitale Medien bedroht die Funktion politischer Öffentlichkeit in Demokratien, die eigentlich kollektives Handeln und Problemlösung ermöglichen soll. Damit entsteht auch der Nährboden für die Verbreitung von Desinformationen. Verschiedene Studien zeigen, dass virale Desinformationen von den oben beschriebenen Mechanismen emotionaler Empörung profitieren (Bakir & McStay, 2018; McLoughlin et al., 2024; Vosoughi et al., 2018), und dass Nutzende weniger kritisch gegenüber Desinformationen sind. Auch teilen sie eher Desinformationen, wenn sie sich auf ihre unmittelbaren emotionalen Reaktionen verlassen, anstatt über den Nachrichteninhalt nachzudenken (Bago et al., 2020; Martel et al., 2020). Die genannte konfliktorientierte und negativ-emotionalisierte Verschiebung politischen Diskurses im digitalen Raum trägt somit zu den Bedingungen bei, in denen sich Desinformationen verbreiten können (W. J. Brady et al., 2023).

Ein weiterer Teilaspekt der oben geschilderten diskursiven Verschiebungen ist der strategische Einsatz dessen, was Quant et al (2022) als "Dark Participation" beschreiben, u.a. unzivilisierte Kommunikation, Provokationen oder eben auch Desinformationen) zum Agenda Setting und Hacking des öffentlichen Diskurses. Ein prominentes Beispiel findet sich im Kontext der US-Präsidentenwahl 2016, als Donald Trump durch den Einsatz entsprechender Stilmittel und Inhalte sowohl bei den Vorwahlen der Republikaner als auch bei der Präsidentenwahl selbst mit seinen digitalen Kanälen einen überproportionalen Anteil der medialen Aufmerksamkeit auf sich zog (Schroeder 2018, Wells et al. 2016). Steve Bannon, damals Trumps Wahlkampfmanager und Mitgründer des rechtsextremen Portals Breitbart News, beschrieb den Grundsatz der dahinterstehenden kommunikativen Strategie mit den Worten: „Flood the zone with shit.“ Entscheidend für den Erfolg ist die erfolgreiche Besetzung der öffentlichen Aufmerksamkeit. Die Inhalte und das Echo (positive oder negative Berichterstattung zum Beispiel) werden zweitrangig. Hierdurch arbeiten sich Medien, politische Konkurrent:innen und Bürger:innen im überproportionalen Ausmaß an Narrativen und Themen ab, die von Provokateuren gesetzt werden. Für andere oder eigene Themen und Diskurse bleibt immer weniger Raum.

Die US-Journalistin Julia Angwin bezeichnet diese Strategie des Agenda Hackings durch Tabubrüche, Provokationen und Desinformationen als "DDoS attack on our minds". DDoS Attacken (Distributed Denial-of-service attack) ist eigentlich ein Begriff aus der Cybersicherheit und beschreibt die Überwältigung der Bereitstellung eines Webservice (z.B. einer Webseite) durch eine Flut an (Fake-)Abrufanfragen. Übertragen auf die Strategie der digitalen Agenda

Hackings durch Stilmittel und Inhalte aus dem Bereich der Dark Participation lässt sich dies als eine systematische Überforderung des politischen Willensbildungsprozesses und der Diskurs- und Informationspraktiken beschreiben, die diesem zugrunde liegen. Diese erfolgt durch ständige, inhaltlich oft mehr oder weniger leere, aber im Agenda-Setting erfolgreiche kommunikative Vorstöße. So wird die Aufmerksamkeitsökonomie erfolgreich gehackt, gleichzeitig werden mit dem Tabubruch transportierte Narrative im Sinne der eigenen politischen Agenda normalisiert.

Eine umfassende Analyse von 496 wissenschaftlichen Studien zum Zusammenhang von digitalen Medien und Demokratie im weltweiten Vergleich zeigt ein gemischtes Bild (Lorenz-Spreen et al., 2023). Die Ergebnisse deuten darauf hin, dass sich die Verbreitung digitaler Medien in Autokratien und aufstrebenden Demokratien positiv auf die politische Beteiligung und den Informationskonsum von Bürger:innen auswirkt. In etablierten Demokratien geht die Verbreitung digitaler Medien mit einem höheren Grad an Populismus, Polarisierung und Vertrauensverlust einher (Lorenz-Spreen et al., 2023). Entsprechend muss das Phänomen der Desinformation in diesem breiteren Kontext betrachtet werden, um mögliche Auswirkungen auf die Gesellschaft bestimmen zu können. Dieser Logik folgend werden zunächst gesellschaftliche Veränderungsprozesse betrachtet, mit dem Ziel, die Verbindung dieser Prozesse mit öffentlichen Diskursen in digitalen Medien sowie Desinformationen darzustellen. Hierbei werden zunächst übergeordnete Phänomene der gesellschaftlichen Spaltung beziehungsweise Polarisierung sowie des Vertrauensverlustes in gesellschaftliche Institutionen betrachtet, bevor auf die Auswirkung von Desinformation auf Funktionsträger:innen und marginalisierte Gruppen eingegangen wird. Den Abschluss bildet der Wissensstand zur Verbindung von Desinformation, digitalem Diskurs sowie Gewalt und Extremismus.

Gesellschaftliche Polarisierung ist ein vielschichtiges Phänomen, das sich in verschiedenen Ausprägungen zeigt und unterschiedliche Auswirkungen auf die Gesellschaft hat. Im wissenschaftlichen Diskurs zu diesem Phänomen wird die Notwendigkeit betont, zwischen diesen verschiedenen Ausprägungen zu unterscheiden, um die Dynamiken und Folgen adäquat erfassen zu können. Hierbei wird unter anderem die (A) ideologische Polarisierung, (B) die affektive Polarisierung, (C) sowie themenbezogene, (D) diskursive und (E) destruktive Polarisierung unterschieden.

(A) Die ideologische Polarisierung beschreibt das zunehmende Auseinanderdriften politischer Positionen, Ideologien und Einstellungen innerhalb einer Gesellschaft. Ein besonderes Augenmerk liegt auf den politischen Eliten¹¹ und Parteien, genauer: auf der Distanz zwischen den politischen und ideologischen Standpunkten von Akteur:innen (Esau et al., 2024). Die wissenschaftlichen Erkenntnisse zur Entwicklung der ideologischen Polarisierung zeigen, dass diese in den letzten Jahrzehnten in vielen westlichen Demokratien zugenommen hat. Dabei ist

¹¹ Der Begriff der politischen Eliten wird in Studien teilweise unterschiedlich eingesetzt und verstanden. Häufig bezieht sich der Begriff auf eine kleinere Gruppe von Individuen, die in einem politischen System über Macht und Einfluss verfügen (z. B. Politiker:innen, hohe politische Beamte, sowie weitere politische Entscheidungsträger:innen).

der Grad der ideologischen Polarisierung in den USA im Vergleich zu Europa deutlich stärker ausgeprägt (Esau 2024; Wagner, 2024). Ein Blick auf Europa zeigt, dass der Grad der ideologischen Polarisierung in Mittel- und Osteuropa tendenziell stärker ausgeprägt ist als in Ländern Nordwesteuropas (Esau 2024; Wagner, 2024). Die Studien zeigen auch, dass der Grad der ideologischen Polarisierung stark mit der Höhe der affektiven Polarisierung zusammenhängt.

(B) Die affektive Polarisierung beschreibt die Feindseligkeit zwischen den Anhänger:innen verschiedener politischer Parteien oder Gruppen. Erfasst wird die Feindseligkeit durch die Einstellungen gegenüber bestimmten Parteien und politischen Gruppen durch Erfassung der Intensität der negativen Gefühle gegenüber dem politischen Gegner (Esau et al., 2024). Die Studien zeigen, dass die affektive Polarisierung in vielen europäischen Mehrparteiensystemen stark ausgeprägt ist (Esau 2024; Wagner, 2024). Hierbei sind insbesondere die Gesellschaften der mittel-, ost-, und südeuropäischen Länder besonders stark betroffen (Esau 2024; Wagner, 2024).

(C) Die themenbezogene Polarisierung bezieht sich auf die Divergenz von Einstellungen und Meinungen zu konkreten politischen Themen oder gesellschaftlichen Problemen (Esau et al., 2024). Dies können unter anderem Migration, Klimawandel, Gesundheitspolitik oder soziale Gerechtigkeit sein (z.B. Meyer et al., 2023). Unterschiede zeigen sich beim Urteil über die Schwere des Problems, über die Verantwortlichkeit für Probleme und über die diskutierten Lösungsstrategien. Die themenbezogene Polarisierung ist in vielen Fällen eng mit der affektiven und ideologischen Polarisierung verknüpft. Die drei genannten Ausprägungen der Polarisierung können sich gegenseitig verstärken (Esau 2024; Wagner, 2024).

(D) Zu nennen ist noch die diskursive Polarisierung. Sie bezeichnet die zunehmende Divergenz und Konfrontation in der öffentlichen Kommunikation. Sie drückt sich in der Art und Weise, wie politische Themen dargestellt werden, aus, in der spezifischen Verwendung von Sprache und Argumentationsmustern, sowie in den Interaktionsmustern zwischen verschiedenen Gruppen in traditionellen und digitalen Medien (Brüggemann & Meyer, 2023).

Die dargestellten Polarisierungs-Konzepte beschreiben ein zunehmendes Auseinanderdriften zwischen Parteien und gesellschaftlichen Gruppen. Inwieweit es sich bei diesem Veränderungsprozess um ein zwangsläufig negatives Phänomen handelt, ist Gegenstand wissenschaftlicher und gesellschaftlicher Debatten. Einige Autor:innen argumentieren, dass es sich bei der gesellschaftlichen Polarisierung um das Ergebnis von Kämpfen für soziale Gerechtigkeit und Gleichheit handelt, die von marginalisierten Gruppen initiiert werden und auf eine Gegenreaktion dominanter Gruppen stoßen (Kreiss & McGregor, 2024). Vor diesem Hintergrund wären einige Formen der Polarisierung notwendiges Übel oder möglicherweise sogar wichtiger Bestandteil von normativ wünschenswerten gesellschaftlichen Veränderungsprozessen.

(E) Das Konzept der destruktiven Polarisierung ermöglicht eine differenzierte Betrachtung von Polarisierung, indem es sich auf Formen der Polarisierung bezieht, die eine Gefahr für demokratische Prozesse und den gesellschaftlichen Zusammenhalt darstellen (Esau et al., 2024). Merkmale hierfür sind unter anderem ein feindseliger und von Misstrauen geprägter öffentlicher Diskurs, der beispielsweise durch illegale Hassrede und Desinformation in sozialen Medien gekennzeichnet ist. Weitere Merkmale sind das Untergraben der Bereitschaft, politische Kompromisse zu schließen, das Infragestellen der Legitimität demokratischer Institutionen und Verfahren, beispielsweise durch die Verbreitung von Falschnachrichten, sowie die Erosion des sozialen Kapitals (Esau et al., 2024), die zu einer zunehmenden Gewaltbereitschaft gegenüber Andersdenkenden führen kann.

Das Zusammenspiel der sozialen Medien, Desinformation sowie der gesellschaftlichen Polarisierung sind Bestandteil eines intensiven öffentlichen Diskurses. Journalist:innen und Politiker:innen wie die ehemaligen Präsidenten Joachim Gauck und Barack Obama haben soziale und digitale Medien als Schuldige für zunehmende gesellschaftliche Polarisierung bezeichnet (Gauck, 2016; Nakamura, 2016). Inwieweit diese gesellschaftlichen Veränderungsprozesse wirklich ausschließlich auf digitale Medien und desinformierende Inhalte zurückzuführen sind, ist dabei unklar. Es gibt jedoch wissenschaftliche Erkenntnisse, die für einen Zusammenhang von digitalen Medien und Formen der Polarisierung sprechen. Denn das Design digitaler Medien und der Einsatz von algorithmischen Empfehlungssystemen kann dazu beitragen, dass Nutzende verstärkt Inhalten ausgesetzt sind, die ihre eigene politische Meinung bestärken und somit auf einer gesamtgesellschaftlichen Ebene zu einer verstärkten Polarisierung beitragen (Arora et al., 2022). Auf diesen Wirkungsweisen basiert häufig die Gestaltung desinformierender Inhalte, die teilweise gezielt an die Meinungen und Einstellungen der anvisierten Zielgruppe angepasst werden. Insofern verfügen desinformierende Inhalte über das Potenzial, zu Polarisierungsprozessen beizutragen.

Ein weiterer Aspekt ist der Vertrauensverlust in demokratische Prozesse und Institutionen. In einer Demokratie kommt Vertrauen eine Doppelrolle zu: Es verbindet Bürger:innen mit politischen Institutionen, ermöglicht kollektive Entscheidungsfindung und verringert die Transaktionskosten politischer Interaktionen. Politisches Vertrauen lässt sich in diesem Zusammenhang als die Erwartung der Bürger:innen charakterisieren, dass politische Akteur:innen und Institutionen in ihrem besten Interesse handeln und grundlegende Normen wie Gerechtigkeit, Transparenz und Effizienz einhalten. Dieses Vertrauen bildet die Grundlage für die Legitimation politischer Systeme. Vertrauen ist ein maßgebliches Element politischer Prozesse und bildet die zentrale Grundlage für das Funktionieren demokratischer Gesellschaften (Levi & Stoker, 2000; Zmerli, 2023). Dies gilt sowohl auf der Ebene des gesellschaftlichen Miteinanders als auch in Bezug auf politische Akteur:innen, staatliche Institutionen und Medien (Sønderskov & Dinesen, 2016). Die Herausbildung von gesellschaftlichem Vertrauen ist Bestandteil der politischen Sozialisation und ergibt sich aus der Summe individueller Erfahrung und gesellschaftlichen Rahmenbedingungen (Delhey & Newton, 2003). Ohne Vertrauen verlieren demokratische Institutionen ihre Handlungsfähigkeit, da Bürger:innen misstrauischer gegenüber Gesetzen und politischen Entscheidungen werden

und seltener bereit sind, Kompromisse einzugehen (Citrin & Stoker, 2018). Auf institutioneller Ebene erleichtert Vertrauen die Funktionsweise komplexer Entscheidungsprozesse. In einer Demokratie, die sich durch eine Vielzahl von Interessen und Meinungen kennzeichnet, ist Vertrauen in Institutionen wie Parlamente, Gerichte und Verwaltungen entscheidend, um politische Stabilität und gesellschaftlichen Zusammenhalt zu gewährleisten. Vertrauen ist auch der Schlüssel für kooperative Beziehungen zwischen Regierungen und ihren Bürgern, insbesondere in akuten Krisen wie etwa der COVID-19-Pandemie, in denen schnelle Maßnahmen erforderlich sind (Siegrist & Cvetkovich, 2000). Die Beziehung zwischen Vertrauen und politischer Partizipation ist dabei aber komplex. So findet sich beispielsweise Evidenz dafür, dass ein Mangel an Vertrauen zwar die Teilhabe an institutionalisierter Politik mindert, nicht aber die Beteiligung an nicht-institutionalisiertem politischen Handeln (Hooghe & Marien, 2013).

Die Digitalisierung hat den öffentlichen Diskurs und die Interaktion zwischen Bürger:innen, politischen Akteur:innen und Institutionen grundlegend verändert. Digitale Plattformen dienen heute als zentrale Orte der politischen Kommunikation und Meinungsbildung. Während Plattformen neue Möglichkeiten zur Teilhabe und Mobilisierung eröffnen, stellen sie zugleich eine erhebliche Herausforderung für das Vertrauen in politische Prozesse dar. Dabei findet sich wenig Evidenz für die Annahme, dass Nutzende unfreiwillig und unwissentlich durch algorithmischen Selektionsprozesse auf digitalen Plattformen beeinflusst werden, sondern mehr Belege dafür, dass die Plattformen genutzt werden, um die eigenen Standpunkte und Interessen zu ergänzen und zu stützen. Dieser Prozess kann zu einer individuellen Radikalisierung führen (Dahlgren, 2021; Stark et al., 2021). Da die sozialen Medien auch für extreme Ansichten einen vielfältigen Marktplatz bieten, auf der sich Gleichgesinnte finden, können sie den Vertrauensverlust gegenüber gesellschaftlichen Institutionen fördern, wenn sie bestehende Zweifel an deren Legitimität verstärken (Zimmermann & Kohring, 2020). Dies trägt zur Fragmentierung des öffentlichen Diskurses bei und erschwert die Bildung eines gemeinsamen Verständnisses politischer Themen (Bright, 2018). Dazu trägt auch die Verbreitung von Desinformation bei. Da Desinformationen und Verschwörungserzählungen durch digitale Plattformen schnell und weitreichend verbreitet werden können, kann so das Vertrauen in traditionelle Informationsquellen wie Medien oder staatliche Institutionen untergraben werden (Ognyanova et al., 2020). Diese Dynamik hat das Potenzial, demokratische Prozesse zu destabilisieren, indem sie Zweifel an der Legitimität von Wahlen oder politischen Entscheidungen schürt. Dabei wird eine verstärkte Nutzung sozialer Medien mit einem reduzierten Vertrauen in traditionelle Nachrichtenquellen in Verbindung gebracht, ohne dass eine kausale Beziehung eindeutig bestimmt werden kann (Karlsen & Aalberg, 2023; Park et al., 2020).

In diesem Wirkungsgefüge nehmen Funktionsträger:innen einer demokratischen Gesellschaft wie Politiker:innen, Amtsträger:innen, Behörden, Medien und Journalist:innen, oder Wissenschaftler:innen eine besondere Rolle ein. Theoretisch bieten digitale Medien diesen Gruppen direkte Kommunikationskanäle zu den Bürger:innen, um beispielsweise kursierenden Desinformationen entgegenzutreten und so mögliche Unsicherheiten, Falschmeldungen und

Informations-Voids schnell beseitigen zu können. Tatsächlich scheint dies in der Praxis eher begrenzt und/oder mit begrenzter Wirkung zu geschehen. Debunking¹² von Falschmeldungen durch etablierte Medien und Journalist:innen erreicht meist deutlich weniger Nutzer:innen als die originale Falschmeldung und/oder kann unabsichtlich zu einer weiteren Verbreitung der Falschmeldung oder der Verschwörungserzählung beitragen (Chan et al., 2017; Schüz & Jones, 2024; Van Der Linden et al., 2021). Zudem fehlen Behörden und anderen Funktionsträger:innen oft die nötige Ausbildung und die nötigen Strukturen und Ressourcen für eine effektive Kommunikation in digitalen Medien. Problematisch ist auch oft die geforderte Schnelligkeit in Fällen, in denen ggf. noch keine gesicherten Informationen verfügbar sind. (Löffelholz et al., 2024; Schulze et al., 2023). Darüber hinaus besteht auch das Risiko, dass diese direkten Kommunikationswege auch von Funktionsträger:innen zur Verbreitung von Desinformationen missbraucht werden, wie beispielsweise durch Donald Trump und seine Administration wiederholt geschehen (Baker, 2024; Washington Post Fact Checker, 2021).

Hinzu kommt, dass diese Funktionsträger:innen regelmäßig Opfer von massiven Angriffen und illegaler Hassrede in digitalen Medien werden, oft verbunden mit Desinformationen, die in diesen Kontexten verbreitet werden. Dies manifestiert sich vielfach in Form von Gewalt, Bevormundung und Sexualisierung, wie etwa in Vergewaltigungsdrohungen, und ist häufig mit analoger Gewalt assoziiert, wie etwa durch physische Gewalt (HateAid et al. 2025). Gerade in besonders politikrelevanten Wissenschaftsbereichen, wie beispielsweise Klimaforschung oder Gesundheitsforschung im Kontext der Corona-Pandemie, aber auch in Feldern wie der Kommunikationswissenschaft berichten Forschende von Online-Anfeindungen (Global Witness, 2023; Nogrady, 2021; Seeger et al., 2024). Auch Politiker:innen und Journalist:innen sind regelmäßig Ziele von Anfeindungen im Netz (Bitschnaus & Eberspach, 2024; Cheruiyot, 2024; Papendick et al., 2020). Dieser Online-Hass kann tiefgreifende Auswirkungen haben. Betroffene berichten von Beeinträchtigungen des psychischen und physischen Wohlbefindens, Stress, Schlaflosigkeit und Depressionen (Cheruiyot, 2024). Die Angriffe können über die berufliche Ebene hinausgehen, wenn sie das Privatleben und Familienmitglieder betreffen. Das kann das allgemeine Wohlbefinden zusätzlich massiv beeinträchtigen (Congress of Local and Regional Authorities, 2022). Besonders schwerwiegend für das demokratische System sind Überlegungen der Betroffenen, die in ihrer Arbeit Selbstzensur zu üben oder sich aus der jeweiligen Tätigkeit zurückzuziehen. Auch mögliche neue Akteur:innen werden potenziell davon abgehalten, entsprechende Tätigkeiten anzustreben (Congress of Local and Regional Authorities, 2022; Seeger et al., 2024). Langfristig bedrohen diese Entwicklungen nicht nur das individuelle Wohl der Betroffenen, sondern auch die Stabilität, Resilienz und Diversität demokratischer Diskurse und Institutionen (Cheruiyot, 2024; Congress of Local and Regional Authorities, 2022).

Zu unterstreichen ist, dass gesamtgesellschaftlich gesehen Hasskommentare im Internet Frauen und marginalisierte Gruppen besonders häufig und in spezifischen Formen treffen. So scheinen Frauen häufiger als Männer und/oder spezifisch mit Verweis auf ihr Geschlecht Ziel von Hasskommentaren zu sein. Auch nehmen diese öfter als bei Männern sexualisierte Formen an

¹² Eine tiefere Betrachtung von Debunking findet sich in Kapitel 5.6.

(Döring & Mohseni, 2019; Hestermann et al., 2021; Jaki & De Smedt, 2019). Neben geschlechtsspezifischen Angriffen sind auch marginalisierte Gruppen stark betroffen: Hass richtet sich überdurchschnittlich häufig gegen Muslim:innen, Jüd:innen, LGBTQIA+-Personen, Sinti:zze und Rom:nja und Menschen mit sichtbarem Migrationshintergrund oder nicht-weiße Menschen. Diese gruppenspezifischen Angriffe scheinen sich in ähnlicher Form auch bei Funktionsträger:innen zu zeigen, wenn beispielsweise Wissenschaftler:innen häufiger Ziel von Online-Hass werden als Wissenschaftler (Seeger et al., 2024). Gesamtgesellschaftlich oft diskriminierte und bei Funktionsträger:innen tendenziell unterrepräsentierte Gruppen werden also besonders oft Zielscheibe von Online-Hass.

Diese politische Gewalt und der gewaltbereite Extremismus hat sich in den letzten Jahren vor allem im Kontext von Rechtsextremismus bzw. -terrorismus sowie in Bezug auf einen verschwörungsideologischen Hintergrund verstärkt (Marwick et al., 2022; Wahlström & Törnberg, 2021). So stellen digitale Medien auch für Akteur:innen und Ideologien mit Gewaltbezug einen digitalen Möglichkeitsraum dar: Ideologien und Narrativen können verbreitet, neue Unterstützer:innen gewonnen werden. Es entsteht ein geschützter Raum zur gemeinsamen Identitätsbildung, der gegenseitigen Bestätigung und fortschreitenden Normverschiebung und Radikalisierung. Und es ist ein Raum für die Bereitstellung praktischer Informationen zum Begehen von politischen Gewalttaten (Gaudette et al., 2022; Karell et al., 2023; Manemann & Reinfrank, 2020; Marwick et al., 2022; Wahlström & Törnberg, 2021). Auch scheint die digitale Konfrontation mit Andersdenkenden Gewaltausbrüche begünstigen zu können (Gallacher et al., 2021). Die Attentäter von Christchurch und Halle (beide 2019) waren eng in rechtsextreme und rechtsterroristische Online-Subkulturen integriert. Die Ausführung der Angriffe war stark von den Symbolen und Codes dieser Milieus beeinflusst, wozu unter anderem die „Gamifizierung“ des Attentats durch einen digitalen Livestream sowie der Einsatz charakteristischer Musik und Handzeichen gehörte (Fielitz & Marcks, 2020; Manemann & Reinfrank, 2020). Zum Hintergrund beider Anschläge lassen sich verschiedene Verschwörungserzählungen wie unter anderem der „Großen Austausch“ zählen. Die „Pizzagate-Schießerei“¹³ resultierte u.a. aus online-kursierenden Verschwörungserzählungen bezüglich einer angeblichen Verwicklung der demokratischen Partei in einen weltweiten Kinderhändlerring (Bleakley, 2023). Aber auch Gewaltausbrüche, die von größeren Gruppen getragen werden, lassen sich mit digitalen Medien, Desinformation und Verschwörungserzählungen verknüpfen. Bekanntheit erlangten beispielsweise die schweren Menschenrechtsverletzungen gegen Rohingya-Muslime 2017 in Myanmar, die von Meta (damals Facebook) durch Untätigkeit und algorithmengestützte Verbreitung von Desinformationen, rassistischer Hassrede, maßgeblich begünstigt wurden, wie Amnesty International berichtete (Amnesty International, 2022). Zu den jüngsten Beispielen zählen die Unruhen in Southporth

¹³ Im US-Wahlkampf wurde die Falschbehauptung verbreitet, Hillary Clinton betreibe ein Pädophilennetzwerk aus einer Pizzeria. Dies führte zu massiven Anfeindungen und Todesdrohungen gegen die Mitarbeiter:innen des Restaurants. Die als „Fake-News-Kampagne“ bezeichnete Kampagne erreichte ihren Höhepunkt, als ein bewaffneter Mann das Lokal stürmte und Schüsse abgab, bevor er festgenommen wurde.

(Institute for Strategic and Dialogue, 2024) oder der Sturm auf das Kapitol 2021 (DFRLab, 2021). Der Informations-Void bezüglich des Hintergrunds der den Unruhen vorausgehenden Messerattacke wurde schnell mit Desinformationen und Falschmeldungen gefüllt, die zu den darauffolgenden schweren Unruhen beigetragen haben (Institute for Strategic Dialogue, 2024). Der Sturm auf das Kapitol war eng mit der lange vorbereiteten #StopTheSteal Desinformationskampagne verknüpft (DFRLab, 2021).

Generell wird in der Forschung jedoch betont, dass gewalttätiger Extremismus und eine vorausgegangene Radikalisierung mit einer Vielzahl von Mechanismen und Prozessen verknüpft sind, von denen digitale Medien einige ermöglichen oder erleichtern können, jedoch maximal einen Teilbereich abdecken (Borum, 2011; Conway, 2016; Marwick et al., 2022). Mit Blick auf die aktuelle Situation in den USA weisen Autor:innen wie Rachel Kleinfeld darauf hin, dass gesamtgesellschaftliche Faktoren wie die ausgeprägte Polarisierung eine wichtige Rolle für den Ausbruch politischer Gewalt spielen (Kleinfeld, 2021). Auch die Bedeutung digitaler Medien für die Verbreitung von Desinformationen sollte nicht überschätzt werden: Zwar wurden Ereignisse wie der Sturm auf das Kapitol unter anderem auch durch digital geprägte Phänomene wie den QAnon Verschwörungsmythos mitgetragen (DFRLab, 2021), die mit dem Sturm verbundene Erzählung von Wahlmanipulation war jedoch Teil der Kampagnenkommunikation (digital und nicht-digital) der Trump Kampagne und wurde – neben der Verbreitung auf Social Media - beispielsweise auch von Mainstream-Sendern wie dem Fernsehsender Fox News verbreitet (Chicago Project on Security & Threats, CPOST, 2022). Dies deckt sich mit Studienergebnissen, die sich in Bezug auf diese Desinformationskampagne Soziale Medien als „zweitrangig“ einordnen und die zentrale Rolle politischer Eliten sowie Mainstreammedien hervorheben (Benkler et al., 2020).



KURZ GEFASST: GESELLSCHAFTLICHE AUSWIRKUNGEN VON DESINFORMATIONEN

Polarisierung: Wissenschaftliche Erkenntnisse deuten auf einen Zusammenhang zwischen digitalen Medien und Formen der Polarisierung hin. Das Design und algorithmische Empfehlungen von Inhalten auf digitalen Plattformen können dazu führen, dass Nutzende verstärkt Inhalte sehen, die ihre eigene politische Meinung unterstützen. Zudem werden Desinformationsinhalte häufig auf die Meinungen der Zielgruppe zugeschnitten und verfügen ebenfalls über das Potential die gesellschaftliche Spaltung zu verstärken.

Vertrauensverlust: Digitale Plattformen bieten ein breites Spektrum an Ansichten, auch Extremen, eine Plattform, auf der sich Gleichgesinnte finden können. Dies kann das Vertrauen in gesellschaftliche Institutionen schwächen, indem bereits bestehende Zweifel

verstärkt werden. Dadurch findet eine Fragmentierung des öffentlichen Diskurses statt, was das Entwickeln eines gesellschaftlich geteilten Verständnisses von politischen Themen erschwert. Diese Entwicklung kann durch Desinformation weiter verschärft werden.

Funktionsträger:innen: Gesellschaftliche und politische Funktionsträger:innen stellen durch ihre (ehrenamtliche) Tätigkeit das Rückgrat einer funktionalen demokratischen Gesellschaft dar. Allerdings werden sie verstärkt das Ziel von politisch motivierten Anfeindungen und Angriffen im digitalen und analogen Raum. Die politische Gewalt und der gewaltbereite Extremismus, insbesondere der Rechtsextremismus sowie bei verschwörungsideologischen Bewegungen, hat sich verstärkt. Digitale Plattformen bieten extremistischen und gewaltorientierten Akteuren einen Raum zur Verbreitung ihrer Ideologie, zur Mobilisierung sowie zur Bereitstellung praktischer Informationen für Gewalttaten. Studien zeigen allerdings, dass digitale Medien nur einen Faktor für extremistische Gewalt darstellen, während weitere gesellschaftliche Faktoren wie die Polarisierung eine zentrale Rolle spielen.

5.2 Wahrnehmung und Bewertung der Gefahren von Desinformationskampagnen und anderen Methoden der Beeinflussung

Um die gesellschaftlichen Auswirkungen von Desinformation im Kontext demokratischer Prozesse umfassend zu analysieren, ist auch die Wahrnehmung von tatsächlichen oder vermeintlichen Desinformationen durch Nutzende und die Gesellschaft zu betrachten. Im Folgenden werden zunächst Studien zur Wahrnehmung und Bewertung von Desinformation in Deutschland und Europa im Ländervergleich vorgestellt, um länderspezifische Einflussfaktoren für die Gefahreinschätzung zu identifizieren. Anschließend werden Studien betrachtet, die individuelle Charakteristika von Nutzenden untersuchen, die sie potenziell anfälliger für Desinformation oder besorgter über deren Auswirkungen machen. Dabei ist zu beachten, dass sich die meisten Studien vorrangig auf Befragungen zur Gefahrenwahrnehmung basieren. Großflächige nicht-experimentelle Studien darüber, wie stark Menschen Desinformation ausgesetzt sind und welchen Effekt dies hat, sind nämlich nur schwer oder gar nicht methodisch valide umsetzbar (siehe Kapitel 5.3.).

Es gibt vier große ländervergleichende Studien zur Risikobewertung von Desinformation durch Nutzer:innen. In einer globalen Studie untersuchen Knuutila et al. (2022) basierend auf Befragungsdaten aus 142 Ländern die Wahrnehmung von Fehlinformation. Die Ergebnisse zeigen zudem erhebliche regionale Unterschiede: In West- und Nordeuropa sahen zum damaligen Zeitpunkt 65 % der Internetnutzenden Fehlinformation als Problem an, während in Osteuropa nur 47 % besorgt waren. Altay (2023) untersucht die Besorgnis über Fehlinformationen in 46 Ländern auf Grundlage des Reuters Digital News Survey (2018–2023). Seine Ergebnisse zeigen, dass diese Besorgnis mit geringer Pressefreiheit und niedrigerem Bruttoinlandsprodukt pro Kopf zusammenhängt. Eine weitere Studie von Vegetti und Mancosu

(2022) in 27 europäischen Ländern ergab, dass Parteipolarisierung und eine niedrige Genauigkeit der Medienberichterstattung zwar nicht mit der wahrgenommenen Exposition gegenüber Fehlinformationen korrelieren, jedoch signifikant mit der allgemeinen Besorgnis über Fehlinformationen.

Während sich diese drei Studien auf Fehlinformationen im Allgemeinen beziehen, unterscheiden Hameleers et al. (2022) zwischen Misinformation (unabsichtlich falsche Informationen) und Desinformation (bewusst verbreitete Falschinformationen). Auch sie konstatieren deutliche Zusammenhänge zwischen der Überzeugung der Gefährlichkeit von Desinformation, der Anfälligkeit für Desinformation und Misinformationen und dem Grad der Pressefreiheit, Vertrauen in Nachrichtenmedien und wahrgenommener Korruption. In Ländern mit einem hohen Grad an Pressefreiheit und Medienvertrauen, wie Dänemark, sind die wahrgenommenen Fehlinformationen und Desinformationen deutlich niedriger als in Ländern mit geringer Pressefreiheit und geringem Medienvertrauen, wie in Ungarn. Diese Zusammenhänge bleiben bestehen, selbst wenn einzelne Einflussfaktoren isoliert betrachtet werden. Die Kategorisierung der Länder nach den systemischen Kriterien Medienvertrauen (Reuters Digital News Report 2021; Newman et al., 2021), Pressefreiheit (Reporters Without Borders, 2020) und wahrgenommene Korruption (Transparency International, 2020). In West- und Nordeuropa, insbesondere in Deutschland, sind Bürger:innen stärker über ungenaue Medienberichterstattung (Misinformatio) besorgt als über gezielt verbreitete Desinformation. In Süd- und Osteuropa hingegen sind beide Wahrnehmungen gleichermaßen stark ausgeprägt.

Eine repräsentative Studie der Bertelsmann Stiftung (Bernhard et al., 2024) mit Befragungsdaten aus dem Jahr 2023 hat neben den USA auch die Wahrnehmung von Desinformationen in Deutschland untersucht. Die Autorinnen und Autoren konstatieren, dass 84 % der Menschen in Deutschland vorsätzlich verbreitete Falschinformationen im Internet als großes oder sehr großes gesellschaftliches Problem ansehen. 81 % betrachten Desinformation als Gefahr für Demokratie und gesellschaftlichen Zusammenhalt. Über 90 % der Befragten glauben, dass die Verbreiter von Desinformation die politische Meinung beeinflussen wollen; 86 % vermuten eine gezielte Manipulation von Wahlergebnissen. Als Hauptquellen von Desinformation wurden Protest- und Aktivistengruppen, Blogger:innen und Influencer:innen, ausländische Regierungen sowie deutsche Politiker:innen und Parteien genannt. Die Studie definiert Desinformation als „falsche Informationen, die absichtlich verbreitet werden, um Schaden anzurichten oder Verunsicherung zu erzeugen“ (Bernhard et al., 2024, S. 17).

Die beschriebenen Studien zeigen, dass die Wahrnehmung von Desinformation stark von politischen und wirtschaftlichen Rahmenbedingungen abhängt. Dabei ist in Ländern mit stark ausgeprägter Pressefreiheit und hohem Medienvertrauen wie Deutschland die Besorgnis im Ländervergleich geringer. Nichtsdestoweniger zeigen die neuesten Zahlen zur Lage in Deutschland, dass eine große Mehrheit der Befragten besorgt ist.

Diese Studierhebungen wurden erstellt, bevor Anfang 2024 publik wurde, dass Russland gezielte Desinformationskampagnen in der EU betreibt, darunter die Verbreitung

propagandistischer Inhalte über die Plattform Voice of Europe, die mittlerweile mit EU-Sanktionen belegt wurde (ARD-aktuell, 2024a). Zudem hat das Auswärtige Amt im Sommer 2024 eine groß angelegte russische Desinformationskampagne mit gefälschten Accounts vermeintlicher europäischer Qualitätsmedien („Doppelgänger-Kampagne“) aufgedeckt, die darauf abzielt, das Vertrauen in demokratische Institutionen zu untergraben und öffentliche Meinungen in Europa zu beeinflussen (ARD-aktuell, 2024b). Es ist unklar, wie sich die Wahrnehmung von Desinformation in Westeuropa (einschließlich Deutschland) durch diese neuen Erkenntnisse verändert hat.

Neben den länderspezifischen Einflussfaktoren prägen auch individuelle Charakteristika, Einstellungen und Nutzungspraktiken die Gefahrenbewertung: Die Analyse der Daten des Reuters Digital News Survey zeigt, dass ältere Menschen, Frauen und Personen mit mindestens einem Bachelor-Abschluss tendenziell besorgter über Fehlinformationen sind (Altay, 2023). Auch die politische Orientierung beeinflusst die Besorgnis: Personen ohne starke politische Zugehörigkeit sind weniger besorgt als Menschen mit starker rechts- oder linksgerichteter sowie zentristischer Haltung. Neben demografischen Faktoren spielen einstellungs- und verhaltensbezogene Aspekte eine Rolle bei der Gefahreinschätzung von Desinformation. Hohlfeld et al. (2020) haben eine nicht bevölkerungsrepräsentative Befragung in Deutschland durchgeführt, in der eine bereits aus anderen Ländern etablierte Third-Person-Dynamik auch in Deutschland gezeigt werden konnte: Obwohl die wenigsten Personen sich selbst als beeinflussbar bezeichnen (first person), glaubt ein Drittel, dass die Wahlabsichten des persönlichen Umfelds (second person), mehr als zwei Drittel sogar, dass die Wahlabsichten der Gesamtbevölkerung (third person) von „Fake News“ beeinflusst werden. Ein Ergebnis, das sich eben auch in der großen Sorge vor Wahlbeeinflussung aus der oben zitierten Bertelsmann-Studie (Bernhard et al., 2024) widerspiegelt.

Im Hinblick auf die eigene Mediennutzung zeigt sich in europäischen Studien, dass eine aktive Auseinandersetzung mit Nachrichten – etwa durch Teilen oder Diskutieren in sozialen Medien – sowie die Nutzung sozialer Medien als Nachrichtenquelle die Besorgnis vor Fehlinformationen erhöhen. Gleichzeitig gehen ein geringeres Interesse an Nachrichten und ein niedrigerer Nachrichtenkonsum mit einer geringeren Besorgnis einher (Almenar et al., 2021; Altay, 2023).

Boulianne und Hoffmann (2024) untersuchen den Zusammenhang zwischen wahrgenommenen Kontakt mit Fehlinformationen auf Facebook und der Besorgnis darüber in Kanada, Frankreich, Großbritannien und den USA. In allen untersuchten Ländern außer Frankreich zeigt die Studie, dass Menschen, die wahrnehmen, Fehlinformationen ausgesetzt zu sein, eine stärkere Besorgnis über Fehlinformationen entwickeln. Da Fehlinformationen stark im politischen Kontext konzentriert sind, steigt die Sorge über Fehlinformationen mit der Menge an politischem Inhalt, die Nutzende auf Facebook konsumieren. Damit replizieren die Autor:innen einen bereits zuvor im Rahmen der COVID-19-Pandemie gefundenen kausalen Zusammenhang von Matthes et al. (2023), die zeigen konnten, dass die wahrgenommene Exposition gegenüber Fehlinformationen bzgl. COVID-19 zu größerer Besorgnis führte. Dabei ist es jedoch wichtig anzuerkennen, dass

die Beziehung zwischen Besorgnis und Exposition wahrscheinlich wechselseitig ist: Menschen, die sich stark um Fehlinformationen sorgen, könnten auch mehr Fehlinformationen wahrnehmen, da sie dafür aufmerksamer oder sensibler sind.

Führungskräfte schätzen die Mis- und Desinformation als kurzfristig größtes Risiko ein. Der Global Risks Report 2024 des Weltwirtschaftsforums analysiert die gravierendsten Risiken, denen die Welt in den kommenden zehn Jahren gegenübersteht. Während langfristig klimabezogene Bedrohungen die Liste der größten Risiken dominieren, identifizieren die befragten Führungskräfte kurzfristig Mis- und Desinformation als das größte globale Risiko (World Economic Forum, 2024). Ähnliche Ergebnisse sind im Bericht für 2025 zu finden (World Economic Forum, 2025). Beide Berichte weisen darauf hin, dass dies insbesondere auf die zunehmende Verbreitung von KI-generierten Inhalten zurückzuführen sei, die falsche Narrative fördern können. Die Expert:innen befürchten, dass Fehlinformationen in den nächsten zwei Jahren die Wahlprozesse in mehreren Volkswirtschaften erheblich beeinträchtigen könnten und spiegeln damit die Besorgnis in der allgemeinen Bevölkerung wider. Zu den befragten Expert:innen gehörten aber wenige Medienforscher:innen und Demokratieexpert:innen, sondern eher Personen im Bereich der Risikoanalyse und der Finanzwirtschaft, was mehr Rückschlüsse über den Ansatz der Studie und die Erwartungen der Führungskräfte zulässt als über die tatsächliche Gefahrenlage.

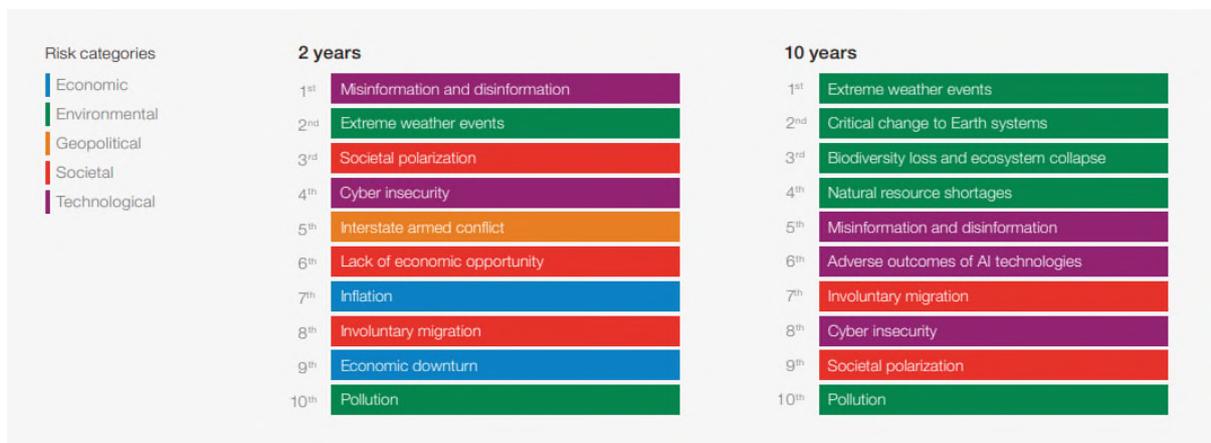


Abb. 1: Priorisierung der globalen Risiken in der Expertenbefragung des Weltwirtschaftsforums 2023-2024

Zusammenfassend lässt sich auf Basis der erwähnten Studien konstatieren, dass die Wahrnehmung von und Besorgnis über Fehlinformationen nicht nur von individuellen Wahrnehmungen und Mediennutzungsverhalten abhängt, sondern auch von strukturellen Bedingungen eines Landes. Länder mit einem stabilen Mediensystem, stark ausgeprägter Pressefreiheit und geringer Korruption bieten ein Umfeld, das das Vertrauen der Bürger:innen stärkt und die negativen Effekte von Fehlinformationen schwächt. Die Entwicklung langfristiger Strategien zur Stärkung von Medienkompetenz und der Resilienz gegenüber Fehlinformationen ist daher eine essenzielle Aufgabe für Politik, Bildung und Gesellschaft.



KURZ GEFASST: WAHRNEHMUNG UND BEWERTUNG VON DESINFORMATION

Sorge vor Desinformation: Die Wahrnehmung von Desinformation variiert je nach politischen und gesellschaftlichen Rahmenbedingungen, wobei Bürgerinnen und Bürger aus Ländern mit hoher Pressefreiheit und Medienvertrauen tendenziell weniger Besorgnis zeigen. Individuelle Faktoren wie Alter, Bildung und politische Orientierung beeinflussen ebenfalls die Einschätzung von Fehlinformationen.

Wechselseitiger Zusammenhang: Studien zeigen, dass ein Zusammenhang zwischen der Wahrnehmung von Desinformation sowie der Sorge über diese besteht: Menschen, die sich stark sorgen, nehmen möglicherweise mehr Desinformation wahr, bzw. Menschen mit mehr Exposition sind besorgter.

5.3 Feststellung und Quantifizierung der Auswirkungen

Für die wissenschaftliche Untersuchung der sozialen, politischen und gesellschaftlichen Auswirkungen von Desinformationskampagnen auf das Individuum, soziale (Teil-)Gruppen sowie die Gesamtgesellschaft wird auf ein breites Repertoire von Werkzeugen und Ansätzen zurückgegriffen. Diese verfügen über spezifische Vor- und Nachteile sowie Einschränkungen. Es besteht kein allumfassender methodischer Ansatz, so dass zielgerichtet geeignete Methoden und Ansätze für die Beantwortung der Teilfragen eingesetzt werden müssen.

Die methodischen Ansätze der Sozialwissenschaften sehen sich mit einigen grundsätzlichen Herausforderungen konfrontiert: (1) dem Problem unscharfer Konzepte, (2) der Schwierigkeit, Kausalität zu erkennen, (3) der potenziellen Dauer und Kleinteiligkeit von Veränderungsprozessen, sowie (4) der Schwierigkeit, adäquate Stichproben zu ziehen.

(1) Die Herausforderung bei unscharfen Konzepten liegt in der Schwierigkeit, eine klare und eindeutige Definition von Begriffen wie der „Desinformation“ vorzunehmen (Tandoc, 2021). Zum einen können diese Begriffe je nach Kontext unterschiedliche Bedeutungen haben. Zum anderen lassen sie sich schwer direkt messen und quantifizieren. Konkret betrifft das unter anderem die Entscheidungen in einem Forschungsprozess zur Bestimmung von Desinformation. Denkbar wäre beispielsweise der Grenzfall eines potenziell satirischen Social Media Posts, der eine Falschbehauptung über eine:n Politiker:in aufstellt. Hier ergeben sich Fragen der Abgrenzbarkeit von Desinformation und Satire. Unklar wären auch deren mögliche Auswirkungen je nach Kontext, auch abhängig davon, ob der/die Leser:in den Social Media Post als Satire erkennt oder nicht.

(2) Der Nachweis von kausalen Zusammenhängen, wie der Einfluss von Desinformationskampagnen auf die politische und gesellschaftliche Entwicklung, ist häufig schwierig. Die Herausforderung liegt hierbei in dem gleichzeitigen Auftreten einer Vielzahl von Einflussfaktoren, in unserem Fall unter anderem die Veränderungen der wirtschaftlichen Lage, der Medienlandschaft und der politischen Kultur. Zum Erreichen eines belastbaren kausalen Zusammenhangs müsste also sichergestellt werden, dass Desinformation die (a) direkte und (b) einzige Ursache für eine (Teil-)Veränderung darstellt. Des Weiteren ist der Nachweis (c) der zeitlichen Abfolge wichtig (Marini & Singer, 1988). In anderen Worten: Zuerst erfolgte das Betrachten von Falschnachrichten und nachfolgend kam es zu einer Veränderung der politischen oder gesellschaftlichen Einstellung. Entsprechend müsste für einen erfolgreichen Nachweis eines kausalen Zusammenhangs in einem Forschungsprojekt auf viele Einflussfaktoren kontrolliert werden, was sowohl unter Real- als auch Laborbedingungen extrem schwierig und teilweise unmöglich ist.

(3) Auch Dauer und Kleinteiligkeit von Veränderungsprozessen stellen die Forschung vor große Herausforderungen. Konkreter: Es handelt sich bei gesellschaftlichen Veränderungen sehr häufig um schleichende, höchst komplexe Prozesse. Das macht den Nachweis von Zusammenhängen komplizierter. Desinformationskampagnen können zum Beispiel erst nach einigen Monaten oder Jahren ihre volle Wirkung entfalten, die dann schwer wissenschaftlich nachweisbar ist.

(4) Forschung braucht eine adäquate Datengrundlage, um belastbare Erkenntnisse liefern zu können. Ist diese Grundlage gegeben, wird von Repräsentativität gesprochen: Merkmale wie Geschlecht, Alter, Einkommen, politische Einstellungen sind in der Datengrundlage (auch Stichprobe genannt) dann gleich verteilt wie in der sogenannten Grundgesamtheit (Diekmann, 2014). In den meisten Fällen bezeichnet die Grundgesamtheit die Bevölkerung eines Landes, sie kann sich aber auch auf kleinere Teilgruppen oder Phänomene beziehen, wie die Nutzenden von bestimmten Social-Media-Plattformen oder Nachrichtenleser:innen. Die Herausforderungen für die sozialwissenschaftliche Forschung liegen hierbei unter anderem in dem Erreichen einer repräsentativen Datengrundlage sowie einer ausreichenden Zahl von Datenpunkten, um auch kleinere gesellschaftliche Teilgruppen abbilden zu können und somit Erkenntnisse über diese Teilgruppen zu gewinnen. Insbesondere diese adäquate Abbildung gestaltet sich in der Forschungspraxis häufig schwierig. Deshalb sind gesellschaftliche Teilgruppen wie beispielsweise Extremist:innen oft nur unzureichend durch gesamtgesellschaftliche Erhebungen abgedeckt. Hierdurch ist die Aussagekraft solcher gesamtgesellschaftlichen Erhebungen in Bezug zu diesen Teilgruppen oft eingeschränkt und es braucht andere methodische Designs um Erkenntnisse gewinnen zu können.

Im Fall der Desinformation sowie der möglichen Auswirkungen auf das Individuum und die Gesellschaft müssen zwei Analyseebenen miteinander verbunden werden. Desinformierende Inhalte müssen erkannt und klassifiziert und deren mögliche Auswirkungen auf Nutzende untersucht werden. Entsprechend erfolgt zunächst die Aufarbeitung von methodischen Ansätzen zur Erkennung und Klassifizierung von Desinformation, gefolgt von einer Darstellung

der Rezeptionsforschungsverfahren, bevor im dritten Teil die wissenschaftliche Analyse der Auswirkung von Desinformation behandelt wird.

Die Analyse von desinformierenden Inhalten konzentriert sich auf das Herausarbeiten von Merkmalen, wie die Art und Weise der Darstellung, der Thematik oder der inhaltlichen Einordnung durch den/die Autor:in. Ziel ist es, Desinformation von vertrauenswürdigen Inhalten zu unterscheiden sowie durch die Inhaltsbeschreibung verschiedene Merkmale herauszuarbeiten. In der Sozial- und Kommunikationswissenschaft wird für solche Zwecke auf die Verfahren der sogenannten Inhaltsanalyse zurückgegriffen. Der breitere Oberbegriff der Inhaltsanalyse umfasst unter anderem (A) Verfahren der qualitativen Inhaltsanalyse, (B) der standardisierten manuellen Verfahren, und (C) Formen der automatisierten Inhaltsanalyse.

(A) Die Verfahren der qualitativen Inhaltsanalyse ermöglichen detailliertere Einblicke in das Phänomen der Desinformation, indem unter anderem Narrative und emotionalisierende Referenzen tiefgehend betrachtet und in einem breiteren Kontext eingeordnet werden können. Das Verfahren fußt auf einer vom Forschenden vorgenommene Klassifizierung der Inhalte in mehrere teilweise zuvor definierte Kategorien. Die können dann im Forschungsprozess erweitert werden (Mayring, 2004).

Der Einsatz von qualitativer Inhaltsanalyse zeichnet sich durch eine große Flexibilität aus, die allerdings einhergeht mit einem höheren Grad an Subjektivität, begrenzter Generalisierbarkeit sowie einem hohen Zeitaufwand. Trotzdem ermöglichen diese Verfahren tiefgreifende Einblicke und tragen zu einem verbesserten Verständnis bei. Das Verfahren kann unterschiedliche Fragestellungen beantworten, wie beispielsweise die Selbstdarstellung Facebooks über die eingeführten Maßnahmen zur Bekämpfung von Desinformation (Iosifidis & Nicoli, 2020).

(B) Standardisierte manuelle Verfahren der Inhaltsanalyse haben dagegen den Vorteil, generalisierbare Aussagen zu ermöglichen (Möhring, 2013). Diese Verfahren weisen allerdings meistens nicht die gleiche Detailtiefe auf wie qualitative Untersuchungen. Für eine standardisierte Inhaltsanalyse wird zunächst ein sogenanntes Kodierbuch erarbeitet. Dieses legt das Vorgehen und die zu identifizierenden Merkmale fest. Danach werden die zu untersuchenden Inhalte durch mehrere Personen anhand der Anleitung des Kodierbuchs kodiert. Die Qualität der Analyse wird anhand der Überprüfung der Kodierentscheidungen zwischen den Personen bestimmt. Stimmt zum Beispiel die Einschätzung der drei kodierenden Personen in 85 % der Fälle überein, wird angenommen, dass es sich bei den untersuchten Inhalten um Desinformation handelt.

Die standardisierte manuelle Inhaltsanalyse ist ein etabliertes Verfahren, das insbesondere für die Analyse etwas längerer Texte, wie in der Studie von Tandoc et al. (2021) über die Merkmale von desinformierenden Nachrichtenartikeln im Vergleich zu Artikeln der Qualitätsmedien gut funktioniert. Grundsätzlich ermöglicht die standardisierte Inhaltsanalyse generalisierbare und reproduzierbare Erkenntnisse. Diese Vorteile gehen allerdings mit einer eingeschränkten

Kontextualisierung, einer begrenzten Flexibilität im Forschungsprozess sowie der Gefahr einher, dass eingesetzte Merkmale zu abstrakt und vereinfachend sind.

Bei einer standardisierten Inhaltsanalyse ist es schwierig, den Kontext angemessen abzubilden. Wenn ein Text zum Beispiel auf ein anderes Bild oder Video verweist, wird dieser zusätzliche Kontext oft nicht erfasst. Des Weiteren ist es aufgrund der standardisierten Vorgehensweise nicht möglich, das Kodierbuch innerhalb des laufenden Forschungsprozesses anzupassen und so auf mögliche neue Entwicklungen zu reagieren. Dies ist insbesondere bei sehr langen Untersuchungszeiträumen (mehreren Monaten) eine Herausforderung. Trotzdem stellt die standardisierte Inhaltsanalyse ein etabliertes und erprobtes Verfahren zur Untersuchung von unterschiedlichsten Medien- und Nachrichteninhalten dar, welche generalisierbare Erkenntnisse über die Merkmale von Desinformation ermöglicht.

(C) Die automatisierte Inhaltsanalyse beschreibt eine Reihe von voll- oder teilautomatisierten Verfahren, welche Text-, Video- oder Audioinhalte klassifizieren können, ohne dass dabei ein Mensch jeden einzelnen Inhalt manuell überprüfen muss. Dadurch ermöglichen die Verfahren der automatisierten Inhaltsanalyse eine sehr effiziente und skalierbare Analyse von großen Datenmengen, die bei einem transparenten Umgang mit dem eingesetzten Verfahren gut zu reproduzieren sind. Die Nachteile dieser Verfahren liegen in der nur bedingten Kontextualisierung von Desinformation, der fehlenden Flexibilität sowie der Gefahr der falschen Einordnung von Inhalten aufgrund von Fehlern der automatisierten Systeme (Heft & Buehling, 2022). Diese Nachteile treffen insbesondere bei Desinformation mit komplexeren Narrationen oder einem großen Maß erforderlichen Kontextwissens zu.

Für automatisierte Verfahren ist es unter anderem schwierig, Inhalte korrekt als Desinformation zu erkennen sowie dazugehörige Merkmale zu identifizieren. Für einige Formen der automatisierten Inhaltsanalyse ist es schwierig zwischen desinformierenden Inhalten und Inhalten, die aufklärend über Desinformation berichten, zu unterscheiden (Heft & Buehling, 2022). Außerdem gibt es plattformspezifische Herausforderungen, die bei der automatisierten Analyse mitgedacht werden müssen, wie beispielsweise deutliche Unterschiede in der Textlänge zwischen kurzen Social-Media-Beiträgen (u.a. auf X) und längeren journalistischen Texten.

Die beschriebenen Verfahren der Inhaltsanalyse lassen sich grob anhand zweier gedanklicher Achsen einordnen. Die erste Achse beschreibt die Anzahl der untersuchten Forschungsobjekte und damit die Vollständigkeit, mit der das Phänomen im Forschungsprozess abgebildet wird. Die zweite Achse stellt die Tiefe der Analyse und möglicher Erkenntnisse dar. Beispielsweise sollen Merkmale von desinformierenden Social Media Posts untersucht werden. Eine qualitative Inhaltsanalyse ermöglicht zwar eine detaillierte Analyse der Merkmale dieser Social-Media-Posts, ist allerdings auch sehr zeitaufwendig. Das führt dazu, dass nur ein kleinerer Teil aller desinformierenden Social-Media-Posts untersucht werden kann. Dieser Nachteil kann durch eine adäquate Stichprobe teilweise ausgeglichen werden. Um bei diesem Beispiel zu bleiben: die standardisierte, manuelle Inhaltsanalyse wäre in der Lage eine deutlich größere Anzahl von

desinformierenden Social-Media-Posts zu untersuchen, bliebe allerdings auf einer etwas abstrakteren Analyse-Ebene. Ähnlich liegt das Problem bei der Anwendung einer automatisierten Inhaltsanalyse. Mit dieser kann zwar eine sehr große Anzahl an Social-Media-Posts analysiert und damit das Phänomen nahezu vollständig abgebildet werden – aber mit dem Nachteil, dass diese Analyse auch oberflächlicher bleibt. Allerdings befinden sich gerade die automatisierten Verfahren mit den Entwicklungen von Large Language Models (LLMs) und anderer AI-Systeme in einer Phase stetiger Verbesserung und Weiterentwicklung. In absehbarer Zeit ist die Entwicklung einer tieferehenden inhaltsanalytischen Untersuchung denkbar.

Inhaltsanalytische Verfahren können auf eine Vielzahl von unterschiedlichen Texten aus verschiedenen Datenquellen, wie journalistischer Berichterstattung oder Social-Media-Posts, angewandt werden. Die Untersuchung von Daten, die auf sogenannten digitalen Spuren von Nutzenden von digitalen Plattformen basiert (Breiter & Hepp, 2018), wird als Social-Media-Analyse bezeichnet. Hierbei werden nicht nur erstellte und gepostete Social-Media Inhalte in den Blick genommen. Vor allem werden die Reaktionen anderer Social-Media-Nutzenden auf diese Inhalte mit einbezogen. Dies ist möglich durch die Auswertung von Social-Media-Metriken, wie die Anzahl der Aufrufe, der Likes oder des Teilens des jeweiligen Beitrags.

Von einer solchen Auswertung profitiert die sogenannte Netzwerkanalyse, die Merkmale eines Social-Media-Beitrags mit der digitalen Reaktion auf diesen in Verbindung zu setzen. Diese weitverbreitete Methode erstellt auf Grundlage von Referenzen (bspw. Reposts und Mentions) der Social-Media-Accounts von Politiker:innen, Medienschaffenden, Einzelpersonen und staatlichen Institutionen ein Netzwerk. Die Analyse der Informationen aus diesem Netzwerk lässt Rückschlüsse über die Relevanz von Akteur:innen, den Informationsfluss sowie den möglichen Einfluss des Netzwerks auf den öffentlichen Diskurs (Heft & Buehling, 2022).

Die Vorteile der Netzwerkanalyse bei der Untersuchung von Social-Media-Daten liegen hier insbesondere in der effizienten und verhältnismäßig schnellen Verarbeitung großer Datenmengen, wodurch der digitale Diskurs auf Social-Media-Plattformen über politische und gesellschaftliche Themen gut dargestellt werden kann (Aïmeur et al., 2023; Heft & Buehling, 2022). Ein besonderer Vorteil liegt hierbei in der Darstellung der Beziehungen zwischen Akteur:innen, die unter anderem durch das Teilen, Kommentieren oder Re-Posten zur Verbreitung von Inhalten, Meinungen und auch Desinformation beitragen. So ist es auch möglich, besonders relevante und zentrale Accounts und dazugehörige Akteur:innen zu identifizieren (Aïmeur et al., 2023; Heft & Buehling, 2022). Neben der Untersuchung von einzelnen Akteur:innen ist es ebenfalls durch die Anwendung von Verfahren der Community-Detection möglich, ganze Gruppen an Accounts mit ähnlichen digitalen Verhaltensmustern zu erkennen (Aïmeur et al., 2023). Dies ermöglicht es, diese Gruppen näher zu betrachten und Aussagen über das interne Gruppenverhalten herauszuarbeiten. Außerdem lässt sich die Rolle dieser Gruppe in Bezug auf die Verbreitung von (Des-)Information, Narrativen und Themensetzung im digitalen Diskurs untersuchen (z.B. Meyer et al., 2023).

Allerdings sind netzwerkanalytische Verfahren in ihrer Aussagekraft auf eine gute Datenqualität angewiesen. Das Erstellen eines guten und umfangreichen Datensatzes ist allerdings aufgrund des teilweise unzureichenden oder ungenügenden Zugangs von Forschenden zu den Daten der großen digitalen Plattformen erschwert. Die Verfahren ermöglichen auch nur eine Aussage über Netzwerkstruktur und die Relevanz von Akteur:innen auf dieser Plattform. Allerdings könnte die Verwendung automatisierter Inhaltsanalysen durch Informationen zu geteilten und online diskutierten Inhalten und Themen ergänzt werden. Eine gewisse Begrenzung ist der Umstand, dass sich die meisten Studien auf die Untersuchung einer spezifischen digitalen Plattform beschränken. So beleuchten sie nur einen Teil des Medien- und Informationssystems. Diese Begrenzung leuchtet insofern ein, dass sich die Verbreitung und Diskussion von politischen Themen auf den drei untersuchten digitalen Plattformen (WhatsApp, Facebook und X) unterscheidet, wie unter anderem die Studie von Yarchi et al. (2021) zeigt.

An dieser Stelle ist anzumerken, dass sich die Erkenntnisse von Social-Media-Analysen auf den Diskurs im öffentlichen, digitalen Raum beziehen und das Ableiten von Rückschlüssen auf die Gesamtbevölkerung problematisch ist. Dies liegt unter anderem daran, dass die analysierten digitalen Verhaltensmuster durch das Verhalten von sogenannten Heavy-Usern getrieben werden (z.B. Hölig, 2018; Kümpel et al., 2015). Diese Heavy-User, die sich überproportional oft digital beteiligen, weisen zu einem erheblichen Teil extremere politische Einstellungen wie starke ideologische Neigungen (Barberá et al., 2015), Parteibindung zu extremen politischen Parteien (Hirndorf, 2020), höhere Medienbias und Feindseligkeit (Tsang & Rojas, 2020) sowie Skepsis gegenüber Medien auf (Yamamoto et al., 2018). Das Phänomen der digitalen Desinformation sowie der sehr parteiische bzw. stark einseitige Medienkonsum scheint besonders stark unter diesen Heavy-Usern ausgeprägt zu sein (Guess, 2021).

Deshalb bietet sich für die Untersuchung der Auswirkungen von Desinformation auf die Nutzenden eine andere Palette an methodischen Vorgehensweisen an. Verschiedene wissenschaftliche Methoden dienen der Erfassung von Einstellungen, Verhaltensweisen und möglichen Einflüssen auf die Verhaltensweisen. Darunter fallen (D) qualitative Interviews, (E) standardisierte Befragungen, sowie (F) automatisierte Beobachtungsverfahren und (G) Feld- und Laborexperimente.

Die genannten Verfahren sind unterschiedlich gut geeignet, stabile Eigenschaften oder leicht veränderliche Zustände sowie Verhaltensweisen von Personen abzubilden. Entsprechend wichtig ist die Unterscheidung in Eigenschaften, welche stabile und langfristige Merkmale einer Person beschreiben, wie beispielsweise der Grad des politischen Interesses oder das Vertrauen in gesellschaftliche Institutionen, und in sogenannte Zustände, welche sich auf kurzfristige und situationsbezogene Stimmung, Aufmerksamkeit oder Motivation einer Person beziehen (Steyer et al., 2015). Im Folgenden werden die genannten methodischen Herangehensweisen auf ihre jeweiligen Vorteile und Grenzen dargestellt, wenn es um die Erfassung der Auswirkung von Desinformation geht.

(D) Ein naheliegender Weg, Einblick in die individuellen Auswirkungen von Desinformation zu gewinnen, ist ein Gespräch mit Nutzenden. Qualitative Interviews thematisieren die Erfahrungen, Einschätzungen und Einstellungen der befragten Person anhand einer offenen oder teilstrukturierten Gesprächsführung. Die Transkription des Gesprächs wird auf wiederkehrende Muster untersucht, anhand derer die tieferen Beweggründe und Mechanismen besser nachvollzogen werden können (Mayring & Fenzl, 2019).

Die Interviewten können auch eigene Verhaltensweisen oder die Nutzung von Desinformation in ihren alltäglichen Leben beschreiben und kontextualisieren, wie die Studie von Gaudette et al. (2022) über die Verzahnung von Rechtsextremismus in der analogen und digitalen Welt beschreibt. Qualitative Interviews erlauben es auch, neue Themen und Aspekte, die möglicherweise in den etablierten und standardisierten Messinstrumenten anderer Verfahren noch nicht abgebildet sind, zu identifizieren.

Erkenntnisse aus qualitativen Interviews lassen sich jedoch aufgrund der geringen Anzahl der Befragten nicht auf eine größere Gruppe oder gar die Gesamtbevölkerung übertragen. Hinzu kommen forschungspraktische Probleme bei der Führung von zwischenmenschlichen Gesprächen, wie beispielsweise das Phänomen der sozialen Erwünschtheit: Befragte Personen erfüllen die wahrgenommene Erwartungshaltung des Gegenübers, anstatt die eigenen Meinungen und Verhaltensweisen zu schildern. Auch gibt es mögliche Verzerrungen in der Erinnerung der befragten Personen. Das trifft insbesondere bei Fragen nach der Quantifizierung der Mediennutzung auf (Araujo et al., 2017; Parry et al., 2021). Schließlich ist es nicht einfach, geeignete Gesprächspartner:innen zu finden, vor allem zum Themenbereich Extremismus. Dennoch geben qualitative Interviews detaillierte Einblicke in Phänomene, die in dieser Granularität mit anderen Methoden kaum zu erreichen sind.

(E) Neben dem direkten Gespräch bietet die standardisierte Form des (Online-)Fragebogens eine weitere Möglichkeit, Einstellungen, Meinungen und weitere Merkmale von Personen mittels Selbstauskunft mit Bezug auf Desinformation zu erfassen. Ein Fragebogen setzt sich hierbei aus einer Reihe von Fragen zusammen, die Rückschlüsse auf direkt messbare und indirekt messbare Merkmale und Konstrukte ermöglichen. Direkt messbare Merkmale stellen beispielsweise das Alter dar, während indirekt messbare Merkmale meist mehrere Facetten aufweisen und entsprechend komplexer abzubilden sind (bspw. das Vertrauen in Medien) (Diekmann, 2014; Weichbold, 2009). Damit auch komplexere Merkmale möglichst gut erfasst werden, wird meist auf eine Kombination aus mehreren Fragen zurückgegriffen, welche jeweils einen Aspekt des komplexen Merkmals abbilden sollen (bspw. das Vertrauen in öffentlich-rechtliche Medien) und im späteren Verlauf durch ihre Kombination einen umfassenderen Blick auf das Merkmal ermöglichen. Entsprechend komplex kann die Entwicklung geeigneter Fragen für die Abbildung indirekter Merkmale sein, weshalb bei einem standardisierten Fragebogen meist auf bereits zuvor entwickelte und in anderen Studien getestete Fragen und Antwortmöglichkeiten zurückgegriffen wird.

Indirekte Merkmale umfassen häufig gesellschaftliche und politische Einstellungen sowie psychologische Aspekte. Diese erfassen zu können, ist ein Vorteil der standardisierten (Online-)Befragung. Durch die Verwendung von bereits in anderen Studien eingesetzten und etablierten Fragen können Ergebnisse aus unterschiedlichen Studien besser miteinander verglichen und in Beziehung zueinander gesetzt werden. Außerdem ermöglichen größer angelegte standardisierte Befragungen eine relativ kosteneffiziente Abbildung der untersuchten Bevölkerungsgruppe und sind gut geeignet, um repräsentative und generalisierbare Ergebnisse zu erreichen. Hinzu kommen Vorteile in Bezug auf die Befragungssituation, da durch die Beantwortung eines vorgegebenen Fragebogens die potenziell verzerrende Einflussnahme durch die interviewende Person minimiert wird (Diekmann, 2014; Weichbold, 2009).

Allerdings ermöglicht die standardisierte Befragung keine so tiefen und kontextualisierten Einblicke in ein Phänomen wie die Methode der qualitativen Interviews. Hinzu kommt, dass es auch hier zu Formen der Antwortverzerrung kommen kann. Das Abbilden der gesehenen, gelesenen oder gehörten desinformierenden Inhalte mittels Selbstauskunft ist allerdings aus mehreren Gründen problematisch: die befragten Personen erinnern sich möglicherweise nicht richtig an die von ihnen rezipierten Inhalte (Parry et al., 2021). Dies führt zu Über- oder Unterschätzung der Dauer und Häufigkeit der Mediennutzung. Im besonderen Fall der Desinformation sowie der Nachrichtennutzung kommt hinzu, dass eine Befragung mittels Selbstauskunft voraussetzt, dass die befragte Person einen Medieninhalt als Desinformation erkennt. Deshalb ist die Erfassung von unbewussten und schleichenden Verhaltensänderungen durch Selbstauskunft schwierig. Trotzdem handelt es sich bei der standardisierten Befragung um ein wertvolles und etabliertes Verfahren, das insbesondere in Kombination mit anderen Verfahren das volle Potenzial entfalten kann (Stier, Breuer, et al., 2020).

(F) Eine Möglichkeit sind hier automatisierte Beobachtungsverfahren, sogenanntes Web- oder Mobile-Tracking (Stier et al., 2021). Diese Verfahren zielen auf die automatisierte Aufzeichnung von Nutzungsverhalten auf Desktop-PCs oder Smartphones ab, indem mittels Browser-Plugins oder mobile Apps geöffnete Webseiten oder Apps mit Einwilligung der Teilnehmenden aufgezeichnet werden. So wird ein sehr granulares und genaues Abbild der tatsächlichen Informationsnutzung und Desinformation erzeugt (Guess et al., 2020; Stier et al., 2021).

Die Vorteile dieser Verfahren liegen vor allem in der genauen Messung der tatsächlichen Mediennutzung. Das Problem der ungenauen Selbstauskünfte in Bezug auf die verwendete Häufigkeit und Dauer von Medien wird damit umgangen (Parry et al., 2021). Durch die hohe Granularität der erhaltenen Daten sind detaillierte Einblicke in das Nutzungsverhalten der Studienteilnehmer möglich, wodurch Dauer und Häufigkeit der Exposition von desinformierenden Inhalten pro Person bestimmt werden kann. Damit ist es möglich, nicht so bekannte Nischen-Medienangebote zu erfassen und in die Analyse einzubeziehen (Stier, Kirkizh, et al., 2020b). Eine vollständigere Abbildung der Desinformation ist somit eher möglich.

Allerdings sind automatisierte Beobachtungsverfahren oft teure und aufwendige Formen der Datenerhebung, die im ersten Schritt nur eingeschränkte Aussagen über die besuchten Inhalte

ermöglichen und häufig auf weitere ergänzende Verfahren angewiesen sind, beispielsweise die bereits dargestellte automatisierte Inhaltsanalyse. Des Weiteren müssen Web-Tracking-Studien häufig hohe Datenschutzaufgaben und Daten-Sicherheitsstandards, aufgrund der Vielzahl und Kleinteiligkeit der erhobenen Daten, erfüllen. Hinzu kommen Herausforderungen in Bezug auf die Teilnehmenden von Tracking-Studien, welche häufig nicht den Merkmalsausprägungen der Gesamtbevölkerung entsprechen und somit das Kriterium der Repräsentativität nur bedingt erfüllen (Stier, Kirkizh, et al., 2020b). Eine weitere Einschränkung der meisten aktuell eingesetzten automatisierten Beobachtungsverfahren ist, dass desinformierende und weitere Inhalte innerhalb von Social-Media Anwendungen häufig nicht erfassbar sind. Allerdings handelt es sich bei den automatisierten Beobachtungsverfahren um ein dynamisches und innovatives Feld, das bereits erste Ansätze zur Aufzeichnung von spezifisch festgelegten Inhalten innerhalb von Social-Media-Anwendungen auf dem Smartphone entwickelt hat (Krieter et al., 2024; Muise et al., 2024).

(G) Experimentelle Forschung ermöglicht die Untersuchung der Auswirkungen von Desinformation auf nutzende Personen, indem einzelne Aspekte verändert werden und die Auswirkung dieser Veränderung auf das Verhalten oder die Einstellung der teilnehmenden Personen erfasst wird (Field, 2003). Grundsätzlich handelt es sich bei Experimenten nicht um eine Methode, sondern um eine Art des Forschungsdesigns, welche die bereits dargestellten Methoden beinhalten kann. Experimente zielen darauf ab, nur einen einzelnen Aspekt zu verändern und alle anderen möglichen Einflussfaktoren dabei stabil zu halten. So kann eine Messung in der Veränderung der Einstellung oder des Verhaltens einer Person auf den zuvor veränderten Aspekt zurückgeführt werden. Hierbei wird zwischen sogenannten Feld- und Laborexperimenten unterschieden. Feldexperimente werden in der natürlichen Umgebung durchgeführt, während Laborexperimente in kontrollierten und künstlichen Umgebungen stattfinden. Ein prominentes Beispiel ist die experimentelle Studie von Nyhan et al. (2023), die in Zusammenarbeit mit Facebook eine Veränderung am Algorithmus vornahm, um die Exposition von gleichgerichteten politischen Inhalten zu reduzieren und die Auswirkung dieser Veränderung auf Polarisierung und Wahlverhalten zu kontrollieren.

Experimentelle Forschung hat den Vorteil, dass durch die kontrollierte Veränderung einzelner Aspekte kausale Zusammenhänge zwischen dem Lesen, Hören und Sehen von Desinformation und der Veränderung in den Einstellungen oder Verhaltensweisen nachweisbar sind (Hameleers et al., 2020). Der Einsatz unterschiedlicher Methoden in der experimentellen Situation, wie etwa Befragungen oder Web-Tracking, ermöglicht es, unterschiedlichste Auswirkungen zu untersuchen, wie Mediennutzungsverhalten, Einstellungen und Wahlentscheidungen. Hinzu kommt, dass experimentelle Forschungsdesigns dazu eingesetzt werden können, um potenzielle Maßnahmen zur Bekämpfung von Desinformation zu testen und somit deren Wirksamkeit einzuschätzen (Hameleers et al., 2020).

Allerdings bringt experimentelle Forschung auch eine Reihe von Herausforderungen mit sich. Das Feldexperiment bietet aufgrund der realitätsnahen Umgebung eine bessere Übertragbarkeit auf eine Realsituation, aber das Kontrollieren der Einflussfaktoren in dieser realen Umgebung

ist problematisch. Diese Kontrolle ist in einem Laborexperiment deutlich einfacher zu handhaben, allerdings ist die Übertragbarkeit der relativ künstlichen Laborsituation auf die Realität eingeschränkter (Field, 2003). Außerdem ist es aus ethischer Sicht nicht möglich, jede Form von theoretisch denkbaren Experimenten wirklich durchzuführen. Beispielsweise wäre es ethisch äußerst problematisch, Studienteilnehmenden Desinformation auszusetzen, um die Veränderung zu beobachten. Studienteilnehmende könnten ihr Verhalten verändern, weil sie wissen, dass sie an einer Studie teilnehmen. Das kann zu Verzerrungen in den Ergebnissen führen. Hinzu kommt, dass viele Experimente über eine begrenzte Stichprobengröße verfügen, wodurch die Generalisierbarkeit der Ergebnisse nicht immer gegeben ist. Studien, wie Nyhan et al. (2023), mit mehreren zehntausenden Teilnehmenden sind sehr selten und meist nur in Kooperation mit Social-Media-Plattformen machbar.

Grundsätzlich stellt experimentelle Forschung aufgrund der Möglichkeit, einzelne Aspekte kontrolliert zu verändern und die Auswirkung zu beobachten, das beste Vorgehen zur Untersuchung der Auswirkungen von Desinformation oder anderer Mediennutzung auf menschliches Verhalten und Einstellungen dar. Feldstudien dieser Art, die unter nahezu realen Bedingungen durchgeführt werden, sind allerdings selten. Gleichzeitig werden gerade diese Studien für die Feststellung von verallgemeinerbaren sowie belastbaren Erkenntnissen über (kausale) Mechanismen und Auswirkungen von Desinformation auf die Gesellschaft dringend benötigt. In der Forschungspraxis sind besagte Feldstudien ressourcenintensiv, komplex und benötigen meist den Zugriff auf Plattform-Daten sowie die Möglichkeit, die Funktionalität oder das Design dieser Plattformen temporär zu verändern. Dadurch ist die Umsetzung solcher Forschungsvorhaben schwierig.



KURZ GEFASST: DESINFORMATION – QUANTIFIZIERUNG – FOLGEN

Grundlegende Herausforderungen: Die Definition und Messung von Desinformation wird durch eine nicht klare Abgrenzbarkeit des Phänomens erschwert. Des Weiteren ist es schwierig, einen direkten Ursachen-Wirkungs-Zusammenhang zwischen Desinformationskampagnen und gesellschaftlichen Veränderungen nachzuweisen, da eine Vielzahl von Einflussfaktoren gleichzeitig vorkommen. Zusätzlich verlaufen gesellschaftliche Veränderungen meist langsam, was den Nachweis weiter erschwert. Hinzu kommt, dass verlässliche und repräsentative Daten benötigt werden, welche allerdings für einige gesellschaftliche Teilgruppen (bspw. Extremisten) schwer zu erheben sind.

Merkmale von Desinformation: Die Untersuchung von Desinformation mit inhaltsanalytischen Verfahren ermöglichen es, Aussagen über die Merkmale von

desinformierenden Inhalten, wie die behandelten Themen, die Tonalität des Beitrags, sowie in diesem verwendete stilistische Mittel sowie Bilder und Videos, zu treffen.

Verbreitung und Interaktion mit Desinformation: Netzwerkanalytische Verfahren ermöglichen es, die Verbreitung von Desinformation sowie relevante Akteure in einem digitalen Informationsnetzwerk zu untersuchen. Auf diese Art und Weise ist es möglich, einzelne oder mehrere zentrale Social-Media-Accounts, welche Desinformation verbreiten, sowie deren Relevanz in einem größeren digitalen Diskurs auf einer oder mehreren digitalen Plattformen zu bestimmen.

Rezeption von Desinformation: Die repräsentative Befragung von Teilnehmenden sowie deren automatisierte Beobachtung, mit deren Zustimmung, ermöglichen es Aussagen über den Anteil der Desinformation im individuellen Nachrichten- und Medienkonsum der Menschen zu treffen.

Wirkung und Mechanismen von Desinformation: Experiment-Studien ermöglichen es, verlässliche Aussagen über die Mechanismen von Desinformation und deren Auswirkungen auf die Gesellschaft zu treffen. Hierbei wird gezielt ein einzelner Einflussfaktor verändert und die daraus resultierende Veränderung beobachtet. Während Experiment-Studien in kleinerem Rahmen gut durchführbar sind, werden diese für größere Gruppen oder gar Gesellschaften sowie für den Nachweis langfristiger Veränderungen sehr komplex und aufwändig.

5.4 (Mit-)Ursächlichkeit von Desinformationen für die genannten Auswirkungen

Inwiefern lässt sich eine Ursächlichkeit bzw. Mitursächlichkeit zwischen Desinformationen und den oben genannten Auswirkungen feststellen? Um diese Frage zu beantworten, bedarf es zunächst eines Blickes auf die Verflechtung von Kommunikation und Medien. Denn die Beziehung zwischen Medien und menschlicher Kommunikation ist von einer komplexen Wechselwirkung geprägt (Couldry & Hepp, 2023; Krotz, 2012). Einerseits gestalten Medien maßgeblich die Art, wie Menschen miteinander in Kontakt treten und Informationen austauschen. Andererseits beeinflussen die Nutzenden durch ihr Verhalten und ihre Präferenzen die Entwicklung und Anwendung dieser Medien. Dieser gegenseitige Einfluss manifestiert sich dabei auf individueller Ebene wie auch in größeren gesellschaftlichen Zusammenhängen. Dabei sind Medien und Kommunikation so eng miteinander verflochten, dass eine Trennung kaum noch möglich ist. Zudem hat die Omnipräsenz medialer Kanäle dazu geführt, dass nahezu jeder Informationsaustausch über irgendeine Form von Medium stattfindet. Diese Verschmelzung hat weitreichende Auswirkungen auf die Art, wie wir interagieren und uns verständigen. So entsteht ein dynamisches Geflecht aus medialen und kommunikativen Praktiken, das unsere moderne Informationsgesellschaft prägt (Watolla, 2024).

Eine lineare Verbindung von der Verbreitung von Desinformationen, etwa dem politischen Erfolg extremer oder autoritärer Akteur:innen, und möglichen Auswirkungen lässt sich kaum feststellen. Die Verzahnung von Kommunikation und anderen kulturellen Praktiken und Medienräumen ist dafür zu stark (Hepp, 2018). Ein Beispiel dieser Wechselwirkung ist die Dynamik von Politikangebot und -nachfrage (Rovira Kaltwasser & Mudde, 2017; Rau 2023): Der Erfolg politischer Ideologien erfordert sowohl ein Angebot dieser Ideologien als auch eine Nachfrage seitens der Bürger:innen. Die digitale Transformation beeinflusst diese Prozesse auf beiden Ebenen.

Politikangebot	Opportunitätsstruktur im (politischen) System (external supply side):		Politische Außenseiter:innen (internal supply side):
	<u>Strukturierung der politischen Landschaft:</u> u. a. Wahlsystem, Parteiensystem, politische Normen und Traditionen, Wählermobilität	<u>Strukturierung der Medienlandschaft:</u> u. a. Mediensysteme, kommerzielle Medien vs. öffentlich-rechtliche Medien, journalistische Normen, Regulierungsgrad	Ideologie, Führungsfiguren, Organisation, Unterstützer:innen
		<u>Digitale Transformation politischer Öffentlichkeit:</u> geschwächte Gatekeeper, Position traditioneller Medien sowie verschärfte Aufmerksamkeitsökonomie als Möglichkeitsraum für politische Außenseiter:innen	<u>Digitale Transformation politischer Öffentlichkeit:</u> ressourcengünstige Möglichkeit, für Außenseiter:innen Unterstützer:innen zu erreichen und zu koordinieren
Politiknachfrage	Grundlegende gesellschaftliche Strukturen, Veränderungsprozesse, Konflikte und Ereignisse		
	Globalisierung; Individualisierung; Diversifikation und Liberalisierung; soziale Ungerechtigkeit; soziale Unsicherheit; individuelle, relative und kollektive Deprivationserfahrungen (Entbehungen); regionale Unterschiede (bspw. Ost vs. West, Stadt vs. Land), Migration; individuelle Einstellungen (bspw. Rassismus, Autoritätsglaube); Terrorismus; Krisen		
	<u>Digitale Transformation politischer Öffentlichkeit:</u> erhöhte Sichtbarkeit und Aktivierung und Veränderung von Einstellungen von und bezüglich relevanter Themen und Ereignisse in der Bevölkerung		

Tab. 1: Die digitale Transformation politischer Öffentlichkeit und ihre Wirkung auf Angebot und Nachfrage von Politik.

Die Angebotsseite setzt sich aus der Opportunitätsstruktur im politischen System (external supply side) und der Selbstorganisation (internal supply side) von politischen Akteur:innen zusammen. In der Opportunitätsstruktur gibt es zwei unterschiedliche Ebenen: die Strukturierung der politischen Landschaft (u.a. Wahlsystem, Parteiensystem, politische Normen und Traditionen, Wähler:innenmobilität) und die Strukturierung der Medienlandschaft (u.a. Mediensysteme, kommerzielle Medien vs. öffentlich-rechtliche Medien, journalistische Normen,

Regulierungsgrad). Insbesondere die Medienlandschaft ist, wie in den oben ausgeführt, von der digitalen Transformation geprägt (geschwächte Gatekeeper-Position traditioneller Medien sowie verschärfte Aufmerksamkeitsökonomie als Möglichkeitsraum für (autoritäre) Akteur:innen jenseits des tradierten politischen Spektrums). Die Selbstorganisation (internal supply side) bezieht sich auf Ideologie, Führungsfiguren, Organisation, Unterstützer:innen. Auch diese wird durch die digitale Transformation nachhaltig verändert, da digitale Medien es möglich machen, ressourcengünstig Unterstützer:innen zu erreichen und zu koordinieren (relevant für Inward Communication; siehe Kaiser & Rauchfleisch 2019).

Damit die angebotene politische Ideologie Erfolg hat, muss sie jedoch von den Bürger:innen nachgefragt werden: Dieser Nachfrage liegen grundlegende gesellschaftliche Strukturen, Veränderungsprozesse, Konflikte und Ereignisse zugrunde, wie Diversifikation und Liberalisierung, soziale Ungerechtigkeit, soziale Unsicherheit, individuelle, relative und kollektive Deprivationserfahrungen, regionale Unterschiede (bspw. Ost vs. West, Stadt vs. Land), Migration, individuelle Einstellungen (bspw. Rassismus, Autoritätsglaube), Terrorismus und Krisen. Was in den Augen der Bevölkerung als wichtiges Problem wahrgenommen wird, kann durch die digitale Transformation politischer Öffentlichkeit mit beeinflusst werden: Digitale Medien können die Sichtbarkeit von Themen und Ereignissen erhöhen, Aktivierung auslösen und Veränderung von Einstellungen bewirken. Dabei bleibt jedoch zu betonen, dass die zentralen Antriebskräfte dieser gesellschaftlichen Nachfrage nach wie vor in den zugrunde liegenden Themen selbst verortet sind. Politische Großthemen, Krisen und grundlegende strukturelle gesellschaftliche Transformationen bestimmen im Wesentlichen die gesellschaftliche Nachfrage nach Politikangeboten (Schroeder, 2018). Insofern lässt sich festhalten: Die politischen Herausforderungen, die unsere Demokratien in den letzten Jahren erleben, haben ihre Wurzeln nicht im digitalen Raum. Vielmehr werden die genannten Themen und Konflikte zunehmend in der digitalen Arena des Internets ausgetragen, geformt und gegebenenfalls verstärkt (Jungherr et al., 2020).

Die genannten Verflechtungen und Wechselwirkungen erschweren es, klare Zusammenhänge oder sogar Kausalitäten zwischen Desinformationen und gesellschaftlichen Auswirkungen empirisch nachzuweisen. Dementsprechend divers ist das Bild, was sich bei der Analyse empirischer Studien, die eben jene Zielsetzung verfolgen, zeigt. In ihrer Studie untersuchen Baqir et al. (2024) die Verbreitung und den Konsum von Nachrichten und Desinformation auf Twitter in Frankreich, Deutschland, Italien und dem Vereinigten Königreich über einen Zeitraum von drei Jahren (2019–2021). Sie unterschieden zwischen vertrauenswürdigen und fragwürdigen Quellen und zeigen, dass trotz des feststellbaren aktiven Konsums fragwürdiger Quellen doch die vertrauenswürdigen Quellen die Online-Diskussion dominieren. Im Gegensatz dazu steht eine Studie zur Beeinflussung der Wahrnehmung von Wahlgerechtigkeit durch Desinformationen (Mauk & Grömping, 2024), in der gezeigt wird, dass Desinformation den Zusammenhang zwischen der tatsächlichen Integrität von Wahlen und der öffentlichen Wahrnehmung dieser Wahlen schwächen. Dies führt zu undifferenzierten und stärker polarisierten Ansichten. In Umgebungen mit einer hohen Dichte an Desinformationen vergrößert sich die

Wahrnehmungskluft zwischen Gewinner:innen und Verlierer:innen erheblich, während diese in Kontexten ohne Desinformationen minimal bleibt.

Was eine mögliche (Mit-)Ursächlichkeit von Desinformationen für gesellschaftliche Auswirkungen angeht, lässt sich also feststellen, dass vor allem die Wahrnehmung gesellschaftlicher Prozesse, Strukturen und Institutionen durch Desinformationen beeinflusst werden¹⁴. Im Kontext von Wahlprozessen zeigen die vorliegenden Studien, dass Wahlprozesse selbst weniger von Desinformationen betroffen sind, sondern sich vor allem die Wahrnehmung von Wahlprozessen verändert.

5.5 Beobachtung neuester Entwicklungen

Betrachtet man die Auswirkungen von Desinformation, die Möglichkeiten, diese zu messen, und die Frage der Ursächlichkeit von Desinformation für gesellschaftliche und politische Veränderungsprozesse im Zusammenhang, können aus den berücksichtigten Studien zwei relevante Entwicklungen der letzten Jahre abgeleitet werden. Erstens eine zunehmende affektive Polarisierung der europäischen und deutschen Gesellschaft. Inwieweit es sich bei dieser um eine normativ wünschenswerte gesellschaftliche Transformation¹⁵ oder um destruktive Polarisierung handelt, ist Bestandteil wissenschaftlicher Debatten und aktuell nicht einzuschätzen. Der zweite wichtige Befund ist die zunehmende Erosion der Vertrauensbasis der Bürger:innen in gesellschaftliche Institutionen und etablierte Medien.

Wie stellen sich nun die jüngsten Entwicklungen dar? Festzustellen sind folgende Tendenzen: (1) die zunehmende Verschiebung hin zu einer fragmentierten Wahrnehmung von politischen Situationen, gesellschaftlichen Problemen und möglicher Lösungen, (2) die zunehmende Etablierung der sozialen Medien als direkter Kommunikationskanal von politischen Akteuren zu potenziellen Wähler:innen, (3) die Aneignung von Bestandteilen der Internetkultur in der politischen Kommunikation, wie zum Beispiel Memes, (4) KI als Verstärker von Desinformation, (5) die Auswirkung tatsächlicher oder von Nutzenden wahrgenommener Desinformation auf Wahlprozesse.

(1) Es zeichnet sich eine Tendenz hin zu einer deutlich stärker ausgeprägten Meinungs- und Wahrnehmungspluralität in der Gesellschaft ab. Diese drückt sich in einer Zersplitterung der Wahrnehmung von politischen Situationen, gesellschaftlichen Problemen und deren Lösungsmöglichkeiten aus. Diese neue Pluralität kann zu einer verbesserten Sichtbarkeit zuvor nicht beachteter gesellschaftlicher Teilgruppen und deren Wahrnehmung führen. Gleichzeitig aber wird es schwieriger, eine gesellschaftlich breit geteilte Sicht auf Kernfragen der Demokratie herzustellen und für die Demokratie notwendige politische Kompromisse zu finden. Das Fehlen

¹⁴ Die Bewertung der Gefährdungspotenziale durch Nutzende werden in Kapitel 5.4 noch genauer beleuchtet.

¹⁵ Eine nähere Erläuterung der wissenschaftlichen Debatte um die Frage, ob es sich bei einer affektiven Polarisierung um einen Bestandteil eines normativ wünschenswerten gesellschaftlichen Transformationsprozess handelt, findet in Kapitel 4.4 statt.

einer geteilten Wahrnehmung der gesamten Gesellschaft stellt einen Nährboden für Desinformationskampagnen dar, die sich die zersplitterte Wahrnehmung zu nutzen machen können.

(2) Hinzu kommt die zunehmende Etablierung der sozialen Medien als direkter Kommunikationskanal von politischen Akteuren zu potenziellen Wähler:innen. Dieser direkte Kommunikationskanal ermöglicht es politischen Akteuren, ihre Meinungen und Ziele direkt einer potenziellen Wählerschaft zu vermitteln. Sie sind nicht mehr angewiesen auf die etablierten Medien, die bisher als Gatekeeper Korrekturen oder Einordnungen der verbreiteten Information liefern konnten (Jungherr et al., 2020). Dieser direkte Kommunikationskanal wird umso relevanter für die politischen Akteure, wenn die potenziellen Wähler:innen wenig oder kein Vertrauen in etablierte Medien haben. Politische Akteure können ihre eigene Darstellung und auch gezielte Desinformation verbreiten und die Einordnung oder Korrektur etablierter Medien mit dem Verweis auf deren fehlende Vertrauenswürdigkeit („Lügenpresse“) diskreditieren. Diese Kommunikationsstrategie lässt sich vor allem bei politisch radikalen bis extremistischen Teilen des Parteispektrums in Deutschland beobachten.

(3) Mit der stärkeren Etablierung der sozialen Medien als Kommunikationskanal findet auch eine Aneignung von Bestandteilen der Internetkultur, wie beispielsweise Memes, in der politischen Kommunikation statt (Lübke, 2024; Ross & Rivers, 2018). Humor als politisches Instrument zu verwenden ist keine neue Entwicklung, allerdings stellt die Verwendung und Adaption von humoristischen Inhalten für die Verbreitung mittels digitaler Medien eine neuere Entwicklung dar (Wagner & Schwarzenegger, 2020). Diese Entwicklung hat zum einen normativ positive Seiten, da diese Form der Kommunikation näher am Medienkonsum jüngerer Generationen ist, die dadurch stärker eingebunden sind. Allerdings weist diese Kommunikationsstrategie auch Risiken für den digitalen Diskurs und die Gesellschaft auf, da unter anderem politische Memes häufig auf eine emotionalisierte Darstellung des politischen Gegners sowie dessen Themen abzielen. Zumindest potenziell könnte das die gesellschaftliche Spaltung verstärken. In Bezug auf die Desinformation stellen humoristische Darstellungsformen, wie Memes, eine besondere Herausforderung dar. Denn sind teilweise schwer oder auch gar nicht von Desinformation abzugrenzen. Während humoristisch zugespitzte und teilweise falsche Darstellungen etablierter Medien als solche erkennbar sind – wenn sie zum Beispiel in einer Satiresendung verbreitet werden –, entfällt diese Möglichkeit aufgrund der häufig unklaren Urheberschaft im digitalen Kontext.

(4) KI und generative KI können die Erstellung von Inhalten, hierunter Desinformation, Deepfakes und Memes, vereinfachen: die technische Qualität falscher Inhalte wird besser, die Quantität höher, es ist ein höherer Grad der Personalisierung möglich. Hinzu kommen die „Halluzinationen“, also das Liefern von plausiblen, aber faktisch falschen oder unsinnigen Informationen, der generativen KI (Xu et al., 2024). Vor dem Hintergrund der schwierigen Messung KI-generierter und -verbreiteter Desinformationskampagnen ist die genaue Quantifizierung komplex, allerdings legen die vorhandenen Erkenntnisse eine Zunahme der von KI-generierten und -verbreiteten

Desinformation nahe. Inwieweit diese allerdings konkrete Auswirkungen auf den politischen und gesellschaftlichen Diskurs und die Meinungsbildung haben, ist unklar (Guess et al., 2023).

(5) Vor diesem Hintergrund besteht die Möglichkeit, dass sich durch Nutzende – vermeintliche wahrgenommene oder tatsächliche – Desinformation auf Wahlprozesse auswirken, ohne, dass es möglich wäre, empirisch fundierte Aussagen über die Wirkung zu machen (Mansell et al., 2025). Ein relevanter Aspekt ist hierbei der steigende Aufwand sowie die Zeit, den politische Kandidat:innen für die Entkräftung von Falschbehauptungen aufwenden müssen. Dies ist insbesondere mit Hinblick auf die Aufmerksamkeitsökonomie problematisch, da die Zeit für das Entkräften von Falschbehauptungen auf Kosten der Platzierung von politischen Themen und mögliche Lösungsvorschläge für gesellschaftliche Probleme in der öffentlichen Debatte geht. Kandidat:innen müssen in ihrem Wahlkampf verstärkt Zeit aufwenden, um zu kommunizieren, was sie nicht sind, als für was sie politisch stehen. Dies könnte insbesondere dann problematisch werden, wenn zu viele Menschen aufgrund ihrer Einstellungen und Überzeugungen Informationen pauschal ablehnen, obwohl diese seriösen Ursprungs sind (Esau et al., 2024).

Ein weiterer Aspekt ist die mögliche Verzahnung von potenzieller oder tatsächlicher Desinformation, die möglicherweise zu einem medialen Skandal führen. Grolleau et al. (2020) argumentieren, dass politische Skandale auch für die betroffenen politischen Akteur:innen durchaus vorteilhaft sind, da diese Aufmerksamkeit generieren, die Möglichkeit zur Selbstinszenierung als Sündenbock oder Opfer bieten, das Ablenken von anderen Problemen ermöglichen und der Benachteiligung von politischen Gegnern dienen kann.

Vor dem Hintergrund der aufgezeichneten Tendenzen lassen sich die Ereignisse um die Jahreswende 2024/2025 kritisch einordnen. So stellt die Präsidentschaftswahl in Rumänien am 24. November 2024 ein Beispiel für die Gefahr der internationalen Einflussnahme und den schwierigen Umgang mit diesen dar. Hier zeigt sich eindrücklich die politische Problematik von Desinformationskampagnen im Wahlkampf. Denn in der Wahl in Rumänien erzielte der als prorussisch geltende Kandidat einen überraschenden Wahlerfolg, woraufhin das Verfassungsgericht die Wahl aufgrund eines russischen hybriden Angriffs in Form einer Desinformationskampagne für ungültig erklärte und eine Wiederholung anordnete. Durch den strategischen Einsatz von inauthentischen und authentischen Social-Media-Accounts wurde versucht, den nationalen Diskurs über relevante und kontroverse Themen zu beeinflussen, so wie es auch bei der Doppelgänger-Kampagne in Deutschland der Fall war (Bayerisches Landesamt für Verfassungsschutz, 2024). Dieser Vorgang offenbart ein politisches Dilemma, da zwar Erkenntnisse über die Existenz einer solchen Desinformationskampagne vorlagen, deren tatsächliche Wirkung zu diesem Zeitpunkt aber unklar war¹⁶. Das Dilemma besteht darin, dass entweder die Wahl eines prorussischen Kandidaten mit Unterstützung einer aus dem Ausland gesteuerten Desinformationskampagne in Kauf genommen werden muss oder dass durch die Annullierung der Wahl das Vertrauen der Bevölkerung in die Demokratie beschädigt wird. Beide Fälle können das Vertrauen in die demokratischen Prozesse und den Glauben an die Souveränität

¹⁶ Zumindest soweit das aus öffentlich zugänglichen Quellen zu beurteilen ist.

des Staates stark beschädigen. Entsprechend ist bei dem Bekanntwerden von Desinformationskampagnen ein überlegter und differenzierter Umgang mit der Existenz und der Auswirkung von Desinformation notwendig, um den gesellschaftlichen Vertrauensverlust einzugrenzen.

Eine weitere Gefahr der internationalen Einflussnahme besteht durch eine zunehmende Verzahnung großer amerikanischer Plattformen mit der US-Regierung. Dies kann zu erhöhtem internationalen Druck auf europäische Akteure führen. Ziel könnte es sein, gegen europäische Nationalstaaten und die EU Deregulierungsbestrebungen durchzusetzen. Inwieweit es sich hierbei um ein realistisches Szenario handelt, ist noch schwer abzuschätzen.

5.6 Begrenzung der Auswirkungen und Handlungsspielräume der Plattformen

Digitale Dienste können auf ihre internen kommunikativen Ökosysteme einwirken, um die Verbreitung von Desinformationen zu reduzieren (Bayer et al., 2021). Besonders Interventionen, die das Vertrauen in verlässliche Informationen stärken und demokratische Meinungsbildungsprozesse fördern, scheinen vielversprechend zu sein. Es gilt zu unterscheiden zwischen Maßnahmen vor der Rezeption von Desinformationen durch Nutzende, Maßnahmen während der Rezeption und Maßnahmen nach der Rezeption.

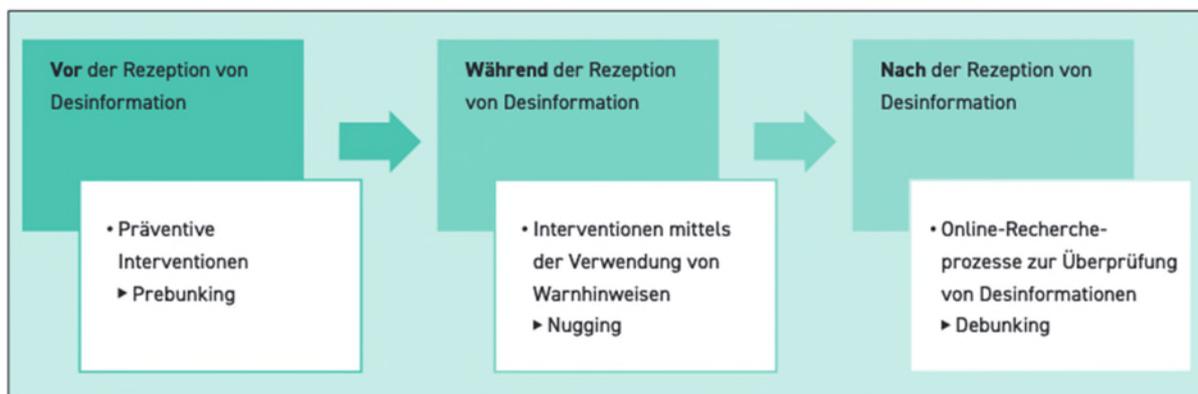


Abb. 2: Interventionen gegen den Glauben an Desinformationen im Rezeptionsprozess (Kessler, 2023, S. 5)

Im Folgenden werden die verschiedenen Möglichkeiten der Intervention vergleichend dargestellt. Sie werden eingesetzt „im Spannungsfeld ihrer Effektivität und Verhältnismäßigkeit,“ (Kessler, 2023, S. 5). Dies ist besonders relevant in Zeiten einfacherer Inhaltsproduktion durch Nutzung von generativer KI (Bontcheva et al., 2024), die sonst gekoppelt mit automatisierten bedingten Verstärkungen („signal boosting“) zu einer Abnahme der Informationsqualität führen kann (Ferrara, 2024).

Zunächst ist zu klären, wie die Handlungsspielräume von Plattformen rechtlich konturiert sind, wenn es um den Kampf gegen Desinformationen geht. Gemäß Art. 34 Abs. 1 i.Vm. Art. 35 DSA besteht eine Pflicht zur Bekämpfung von Desinformation für sehr große Online-Plattformen und Online-Suchmaschinen (Bovermann, 2025). Diese Pflicht lässt sich aber nicht aus dem Normtext

selbst ableiten, sondern nur über die Erwägungsgründe. Art. 35 Abs. 1 und Art. 34 Abs. 1 Unterabs. 2 lit. c DSA verweisen auf die Verpflichtung von Plattformen, systemische Risiken für gesellschaftliche Debatten zu reduzieren; zu diesen systemischen Risiken kann die Verbreitung von Desinformationen gehören. Die Erwägungsgründe tragen diese Interpretation auch. Erwägungsgründe Nr. 2, 83, 84, 88, 95, 106 und insb. 104, stellen zur Genüge klar, dass Desinformationsphänomene vom Ordnungsgeber als demokratiegefährdend gesehen werden. Daher lässt sich für die VLOPs und VLOPSEs schon alleine auf Grundlage Art. 34 f. DSA im Prinzip eine Verpflichtung zur Bekämpfung von Desinformation herleiten. Inhaltlich ist die Pflicht indes nicht im Sekundärrecht konturiert.

Konkretisierend kann hier der „Strengthened Code of Practice on Disinformation“ (Europäische Kommission, 2022) herangezogen werden. Dieser wurde unter Mitwirkung der zentralen VLOPs ausgearbeitet; X (ehemals Twitter) hat sich jedoch nach der Übernahme durch Elon Musk zurückgezogen. Dieser Kodex wird 2025 in einen Code of Conduct im Sinne des Art. 45 DSA weiterentwickelt (Brogi & De Gregorio, 2024). Dieser Verhaltenskodex zur Bekämpfung von Desinformation (Code of Practice on Disinformation) stellt ein innovatives Instrument dar, das erstmals im Jahr 2018 von wichtigen Akteuren der Branche verabschiedet wurde. Ziel war es, selbstregulative Standards zur Eindämmung von Desinformation zu etablieren. Im Juni 2021 begann der Überarbeitungsprozess, der am 16. Juni 2022 mit der Unterzeichnung und Vorstellung des überarbeiteten Kodex abgeschlossen wurde. Dieser Kodex wird künftig Teil eines umfassenderen rechtlichen Rahmens sein, der unter anderem die Gesetzgebung zur Transparenz und Ausrichtung politischer Werbung sowie den DSA umfasst. Für Unterzeichner, die als VLOPs eingestuft werden, soll der Kodex als Maßnahme zur Risikominderung und als Verhaltenskodex im Rahmen der ko-regulatorischen Struktur des DSA anerkannt werden. Der überarbeitete Verhaltenskodex umfasst 44 Verpflichtungen und 128 spezifische Maßnahmen, die sich auf folgende Hauptbereiche erstrecken: Demonetisierung von Desinformation (also Entzug der Möglichkeit, durch Interaktionen Gewinne zu erwirtschaften); Transparenz politischer Werbung; Sicherstellung der Integrität von Diensten; Stärkung der Nutzer mit dem Ziel der Erkennung von Desinformation und Zugang zu vertrauenswürdigen Informationsquellen sowie Maßnahmen zur Förderung der Medienkompetenz und Transparenz bei Empfehlungsalgorithmen und deren Anpassung zur Eindämmung der Verbreitung von Desinformation; Unterstützung der Forschung (auch durch Zugang zu anonymisierte, aggregierte oder öffentlich zugängliche Daten); maßgeblich auch die Stärkung der Faktenprüfer:innen und verstärkte Einbindung in Dienste inklusive einer fairen Vergütung für deren Arbeit); verstärktes Monitoring und insb. Zusammenarbeit mit allen Stakeholdern in Wahlkontexten und Wahlkämpfen.

Der Ansatz des Code of Practice ist es, Selbstverpflichtungen der Plattformen in einem ko-regulativen Ansatz und unter Betonung von Freiwilligkeit und Kooperation vorzusehen. Nun ist nach Art. 45 Abs. 1 DSA die Unterzeichnung solcher Verhaltenskodizes grundsätzlich freiwillig, allerdings wird diese Freiwilligkeit normativ de facto aufgehoben. Erwägungsgrund Nr. 104 deutet dies schon an: Negative Auswirkungen systemischer Risiken auf „Gesellschaft und Demokratie, etwa aufgrund von Desinformation oder manipulativen und missbräuchlichen

Tätigkeiten“, seinen ein paradigmatisches Thema für Codes of Conduct, dazu zählen “koordinierte Tätigkeiten zur Verstärkung von Informationen einschließlich Desinformation, etwa durch Nutzung von Bots oder Scheinkonten für die Erstellung vorsätzlich unrichtiger oder irreführender Informationen“. In diesen Bereichen könnte die Beteiligung von VLOPs an Codes of Conducts und dessen Einhaltung als geeignete Maßnahmen zur Risikominderung angesehen werden. Die Teilnahme ist de facto verpflichtend, denn weigere sich ein VLOP „ohne angemessene Begründung“ an der Anwendung eines solchen Verhaltenskodex zu beteiligen, könnte dies „hinsichtlich der Feststellung möglicher Zuwiderhandlungen (...) im Rahmen dieser Verordnung berücksichtigt werden“. Auch in Art. 75 Abs. 2 DSA wird festgelegt, dass die Kommission bei der Erstellung von Aktionsplänen die Teilnahme an Selbstverpflichtungen berücksichtigen bzw. empfehlen kann.

Auch die DSA Election Guidelines der EU KOM vom April 2024 geben die den VLOPs/VLOSEs Best Practices bei Wahlen vor, hierunter auch mit Blick auf Desinformation. In diesem Maßnahmenkatalog ist unter anderem das frühzeitige Erkennen, Markieren und/oder Stoppen der Weiterverbreitung von Desinformation durch die Plattformen enthalten. Dies kann durch ein engmaschiges Monitoring von Seiten der Plattformen erfolgen. Möglichst frühzeitig können diese kursierende desinformierende Inhalte identifizieren und markieren/labeln und/oder deren Verbreitung einschränken: durch Begrenzung der Sichtbarkeit bzw. Reichweite oder der Teilen- und Weiterleitungsfunktionen. Dabei wird zwischen strafrechtlich relevanten und problematischen, aber legalen Inhalten ("awful but lawful") unterschieden. Ein Fokus liegt auf der Governance von Infrastrukturanbietern und Cross-Plattform-Aktivitäten, die als besonders kritisch für die Sichtbarkeitsreduktion identifiziert wurden (Rau et al., 2022). Hierbei sind die größten Herausforderungen bei der Umsetzung die korrekte Einschätzung der potenziell desinformierenden Inhalte und die notwendige Transparenz beim Ergreifen einschränkender Maßnahmen (Rau et al., 2022; Kessler, 2023).

Im Rahmen der Einschränkung potenzieller Desinformation sind digitale Plattformen neben der selbständigen Analyse auch auf die Unterstützung von Nutzenden angewiesen, die mittels plattform-interner Meldesysteme die Möglichkeit haben, nicht-vertrauenswürdig erscheinende Inhalte an die Content-Moderation der jeweiligen Plattform zu melden. Im Zuge dessen wäre eine Verbesserung der Meldesysteme für eine einfachere und nutzerfreundliche Handhabung ein verhältnismäßig einfaches Mittel (Kessler, 2023).

Neben der Einschränkung der Verbreitungsmöglichkeiten von Desinformation ist die Förderung einer Verbreitung faktenbasierter Informationen ein zweiter Ansatz. Plattformen könnten beispielsweise durch eine Anpassung der algorithmischen Empfehlungssysteme dazu angehalten werden, Inhalte aus vertrauenswürdigen Quellen in ihrer Sichtbarkeit und Verbreitung zu bevorzugen (Rau, 2022; Kessler, 2023). Die Einstufung vertrauenswürdiger Quellen stellt bei diesem Ansatz die größte Herausforderung dar. Zur Sicherstellung einer demokratischen Öffentlichkeit könnte der Prozess der Einstufung von vertrauenswürdigen Quellen nicht einem Privatunternehmen überlassen werden, sondern durch ein demokratisches und staatsfernes Instrument sichergestellt werden. Initiativen wie NewsGuard, ein Angebot von

Microsoft, dienen der Nachrichtenglaubwürdigkeitszertifizierung, sind aber auch privat getragen.

Auch die Nutzung von KI zur Unterstützung der Faktenprüfung ist ein Ansatz, um KI produktiv für die Bekämpfung von Desinformationen einzusetzen. KI-gestützte Systeme können Inhalte mit Datenbanken und vertrauenswürdigen Nachrichtenquellen abgleichen, um falsche Behauptungen schnell und skalierbar zu identifizieren und sollen helfen, die Arbeit von Faktenprüfer:innen erheblich zu erleichtern; solche Modelle sind jedoch mit verschiedenen Schwierigkeiten konfrontiert und erfordern stetige Weiterentwicklungen (Dierickx & Lindén, 2023; Guo et al., 2022). Journalistische Akteur:innen können dies nicht nur für die Verifikation, sondern auch für ihre eigene Textproduktion nutzen, zeigen sich aber dennoch teils ambivalent gegenüber Factchecking-Angeboten (Brandtzaeg et al., 2018). Gleichzeitig können KI-Modelle zwar bei der Moderation von Inhalten in sozialen Netzwerken helfen, indem sie verdächtige Beiträge automatisch erkennen, markieren oder für eine manuelle Überprüfung priorisieren. Es gibt aber verschiedene (ethische) Bedenken gegenüber solchen Einsatzmöglichkeiten (Jungherr, 2023). Algorithmen ermöglichen zudem, vertrauenswürdige Inhalte höher zu ranken, während potenziell problematische Inhalte herabgestuft werden. So lässt sich die Interaktion mit falschen Inhalten signifikant verringern (Gillespie, 2022; Kaiser & Mayer, 2023) sowie fragwürdige Inhalte mit Warnhinweisen oder Labels versehen, um Nutzende auf die Herkunft und Glaubwürdigkeit von Informationen hinzuweisen oder sie an geprüfte, vertrauenswürdige Quellen zu verweisen (Roozenbeek et al., 2023).

Die Erkennung von Bots und koordiniertem Verhalten auf sozialen Netzwerken ist ein weiteres Feld, in dem KI wirksam eingesetzt werden kann: Durch die Identifizierung und Entfernung automatisierter inauthentischer Accounts, die oft zur Verbreitung von Desinformation genutzt werden, können deren Aktivitäten effektiv eingedämmt werden. Durch die fortlaufende Weiterentwicklung von Bots sieht sich jedoch auch dieses Forschungsfeld mit immer neuen Problemen konfrontiert und ist daher ebenfalls zur stetigen Weiterentwicklung geeigneter Erkennungsmethoden herausgefordert (Ferrara, 2023). Gleichzeitig kann KI vertrauenswürdige Inhalte fördern, indem sie Gegendarstellungen generiert und gezielt verbreitet, um falsche Informationen zu widerlegen (Pedrazzi & Oehmer, 2023). Trotz aller Chancen, die sich durch einen solchen Einsatz von KI bieten, darf nicht außer Acht gelassen werden, dass selbst der positiv intentionierte Einsatz von KI-Anwendungen zu zahlreichen unbeabsichtigten, aber dennoch schädlichen Konsequenzen führen kann (DeVerna et al., 2023; Jungherr, 2023; Pedrazzi & Oehmer, 2023).

Ein weiteres potenzielles Instrument ist das sogenannte Prebunking: die präventive Aufklärung der Nutzenden über Desinformation. Dies könnte beispielsweise durch das Einfügen von generischen oder spezifischen Aufklärungstexten innerhalb des Feeds der jeweiligen Plattform geschehen (Kessler, 2023). Dieser Ansatz könnte allerdings auch unerwünschte Effekte mit sich ziehen: Nutzenden würde ein übertriebenes Maß von Desinformation suggeriert, was die gesellschaftliche Vertrauensbasis schwächen könnte, wie oben gezeigt.

In eine ähnliche Richtung geht die Verwendung von Nudging mittels Warnhinweisen. Hier können potenziell falsche oder anderweitig desinformierende Inhalte mit Warnhinweisen versehen werden, um Nutzende zu einer kritischen Prüfung des jeweiligen Inhalts zu animieren (Rau, 2022; Kessler, 2023). Das Nudging durch Warnhinweise bringt sehr ähnliche Herausforderungen mit sich wie die Einschränkung der Verbreitung und Sichtbarkeit von Desinformation. Hier kann auch KI als Potenzial verstanden und eingesetzt werden, um automatisch Desinformation zu erkennen und Transparenz herzustellen: mit Hilfe von Natural Language Processing können KI-Modelle Texte bezüglich sprachlicher Muster, widersprüchlicher Aussagen oder fehlender Quellenangaben analysieren (Tajrian et al., 2023). Darüber hinaus ermöglicht die Bild- und Videoanalyse durch KI die Identifikation von Manipulationen, wie beispielsweise Deepfakes, indem Artefakte, Unregelmäßigkeiten oder abweichende Gesichtszüge erkannt werden (Ghai et al., 2024; Patel et al., 2023). KI-gestützte Tools analysieren dabei feine Details wie Augenbewegungen oder Lippsynchronisation, um manipulierte Inhalte zu erkennen (Javed et al., 2024; Kharel et al., 2023). Zusätzlich können Systeme zur Metadaten-Verifizierung oder Wasserzeichen-Technologien dabei helfen, authentische Inhalte klar zu kennzeichnen (Nadimpalli & Rattani, 2024). Obwohl verschiedene Wissenschaftler:innen die Kennzeichnung von Inhalten als Maßnahme zur Wiederherstellung des Vertrauens in authentische Inhalte und gegen die Verbreitung von Desinformation durch Deepfakes vorschlagen (Lee & Shin, 2022), sollten daher die tatsächlichen Auswirkungen auf das Medienvertrauen kritisch hinterfragt und weiter empirisch untersucht werden.

Um Desinformationen wirksam zu begegnen, wird jedoch ein Zusammenspiel mehrerer Maßnahmen von Seiten der Plattformen nötig sein (Kozyreva et al., 2024; Fazio et al., 2024). Erforderlich sind auch weitere Schritte, die digitale Plattformen unternehmen können – jenseits der Maßnahmen innerhalb ihrer Plattformen –, um der Verbreitung von Desinformationen entgegenzuwirken. Dazu gehören unter anderem Transparenzmaßnahmen, etwa durch Transparenzberichte und Forschungszugänge zu Plattformdaten, um Governance-Entscheidungen kritisierbar und hinterfragbar zu machen, sowie durch die Förderung pro-demokratischer Narrative, die den Zusammenhalt fördern. Demokratische Resilienz sollte durch die Förderung von Medien- und Politikbildung sowie die Einbindung demokratischer Akteure in die Gestaltung und Entwicklung von Plattformregeln gestärkt werden.

Ein wesentlicher Aspekt erfolgreicher Bekämpfung von Desinformationen könnte auch ein Fokus auf die strukturellen Bedingungen sein, in denen sich Desinformationen und andere problematische Inhalte besonders erfolgreich verbreiten. Wie bereits dargestellt, gehört hierzu die konfliktorientierte und negativ-emotionalisierte Verschiebung des öffentlichen Diskurses. Die Etablierung alternativer Mechanismen zur Verteilung von Aufmerksamkeit auf digitalen Plattformen könnte diesen Herausforderungen begegnen. Dazu gehört etwa die erwähnte Förderung der Sichtbarkeit von Kommunikator:innen, die sich demokratischen Normen wie faktenbasierter oder ziviler Kommunikation verpflichtet fühlen, durch eine entsprechende Anpassung der algorithmischen Empfehlungssysteme.

Die Stärkung von Plattformen als demokratische Foren erscheint in einer gesellschaftsübergreifenden Perspektive als besonders erfolgversprechend. Plattformen könnten über die Risikoeinschätzungs- und -minderungspflichten verpflichtet werden, Freiheit, Pluralismus und gesellschaftliche Selbstverständigung zu fördern, indem deren Nicht-Präsenz als Risiko gilt (Broughton Micova & Schnurr, 2024). Dies rechtfertigt, Inhalte und Aktivitäten gezielt zu stärken oder einzuschränken, ohne das demokratische Potenzial digitaler Kommunikation zu gefährden. Eine breite Einbindung staatlicher, privater und zivilgesellschaftlicher Akteure soll Machtkonzentration vermeiden und Resilienz fördern (Humprecht et al., 2020). Kartellrecht und hybride Governance-Strukturen, auch sinnvoll gestalteten Regularien und Praxen des Umgangs mit social bots (Makhortykh et al., 2022), also KI-Systeme, die den Anschein erwecken sollen, als würden reale Personen kommunizieren (Gehl & Bakardjieva, 2016; Hepp, 2021), könnten diesen Prozess unterstützen. Die systematische Einbeziehung vielfältiger Akteure stärkt pluralistische Perspektiven (Ranaivoson et al., 2022) und demokratische Resilienz, z.B. durch Deradikalisierungsangebote oder den Aufbau eines „demokratischen Immunsystems“. Prozedurale Garantien wie Transparenzpflichten, Benachrichtigungen und Gegendarstellungsrechte sichern den Grundrechtsschutz. Plattformräte wie das Facebook Oversight Board können ergänzend zu Gerichten wirken (Kettemann & Schulz, 2023). Der DSA sieht die Teilnahme von Online-Plattformen an außergerichtlicher Streitbeilegung durch zertifizierte Streitbeilegungsstellen vor (Ruscheimer et al., 2024). Mittlerweile sind einige außergerichtliche Streitbeilegungsstellen durch verschiedene nationale Digital Services Coordinators (DSC) zertifiziert worden¹⁷.

¹⁷ Liste aller in der EU gemäß DSA zugelassenen Streitbeilegungsstellen: <https://digital-strategy.ec.europa.eu/en/policies/dsa-out-court-dispute-settlement#ecl-inpage-List-of-bodies>

6 FAZIT UND AUSBLICK

Die Verbreitung von Desinformationen auf digitalen Plattformen erfolgt in einem komplexen Zusammenspiel technischer, ökonomischer, rechtlicher, sozialer und psychologischer Mechanismen. Dabei spielen eine Vielfalt an inländischen und ausländischen Hauptakteur:innen bei der gezielten Verbreitung von Desinformationen eine zentrale Rolle: im Wahlkontext insbesondere aus politischen Interessen, in anderen Kontexten zumeist mit der Absicht, „Traffic“ mit monetären Zielen zu generieren. Hierbei handelt es sich um staatliche wie um nicht-staatliche Akteur:innen, darunter ausländische Regierungen und deren Geheimdienste, politische Parteien, Aktivist:innen und kommerzielle Akteur:innen. Ziel von politisch motivierten Desinformationskampagnen ist es, demokratische Prozesse zu destabilisieren und das Vertrauen der europäischen Bevölkerung in nationale und europäische politische Prozesse und Systeme zu schmälern, Unterstützung für politische Maßnahmen der Regierungen zu untergraben sowie Zukunftsangst zu schüren, während kommerzielle Akteur:innen die aufmerksamkeitsökonomischen Prinzipien nutzen, um wirtschaftliche Gewinne zu erzielen.

Technische Mechanismen wie Empfehlungsalgorithmen und koordiniertes Online-Verhalten können die Reichweite von Desinformationen verstärken, und Algorithmen können polarisierende und emotionalisierende Inhalte priorisieren, unabhängig von deren Richtigkeit. Zudem können Suchmaschinen durch manipulative Suchmaschinenoptimierung ebenfalls zur künstlichen gesteigerten Sichtbarkeit irreführender Inhalte beitragen. Das Zusammentreffen von technischen Mechanismen mit sozialen und psychologischen Faktoren schafft eine komplexe Umgebung, die die Verbreitung von Desinformationen durch algorithmische Systeme erleichtert. Dennoch lässt es sich empirisch nicht nachweisen, dass Desinformation eine substanzielle Wirkung auf individuelle oder gesellschaftliche Meinungsbildungsprozesse hat. Vielmehr könnte es sein, dass der verstärkte Fokus auf Desinformation deren Wirkung vergrößert, auch getrieben von Medien, politischen Akteuren und Teilen der Zivilgesellschaft, alle mit lauterer Motiven. Diese Muster sieht man häufig, wenn etwa Statements auf sozialen Medien mit geringerem Publikum in Qualitätsmedien skandalisiert werden und erst dann, quasi nach einer „Nobilitierung“, richtig wirksam werden.

Die Rolle von Künstlicher Intelligenz in diesem Zusammenspiel ist ambivalent: Einerseits wird KI zur Produktion und Verbreitung von Desinformationen wie Deepfakes genutzt, die zunehmend realistischer und schwerer erkennbar werden. Andererseits bietet KI auch Potenziale, Desinformation durch automatische Erkennung manipulierter Inhalte und Unterstützung bei der Faktenprüfung zu erkennen. Zwar nutzt derzeit nur ein kleiner Teil der Bevölkerung dezidierte KI-Anwendungen, um sich aktiv über politische Inhalte zu informieren, doch ist die Verwendung von KI-Anwendungen zur Produktion politischer Informationen bereits angekommen. Zudem muss berücksichtigt werden, dass KI-generierte Inhalte zu politischen Sachverhalten auch auf anderen Wegen an die Öffentlichkeit gelangen und entsprechend von Menschen aufgenommen werden können, beispielsweise über soziale Medien in Form von KI-generierter Desinformation.

Dass Desinformationsakteur:innen generative KI nutzen, liegt auf der Hand. Hinzu kommt die unhinterfragte bzw. -kontrollierte Art und Weise, wie Suchmaschinen verschiedenen Nutzende für identische Abfragen unterschiedliche Quellenreihenfolgen präsentieren, was Selektionsprozesse beeinflussen kann. Diese Randomisierung kann den Einfluss irreführender Quellen verstärken, wenn sie in bestimmten Fällen eine günstige Positionierung erhalten. Es stellt auch erhebliche Herausforderungen für Forschende dar, die versuchen, die Auswirkungen algorithmischer Systeme zu untersuchen. Denn die dynamische Natur der Suchergebnisse erschwert die Bemühungen, verallgemeinerbare Schlussfolgerungen zu ziehen. Die digitale Aufmerksamkeitsökonomie verstärkt diese Problematik, da emotionalisierte und konfliktorientierte Inhalte bevorzugt werden. Das erschwert sachliche Debatten. Aus einer normativen Perspektive ist jedoch eine inhaltliche Vielfalt gegenüber einer (Hyper)Personalisierung systemisch der Vorzug zu geben, wenn ein ausgewogener Medienkonsum angestrebt wird (Verhulst, 2023).

Obwohl Desinformation ein reales Problem ist, zeigen die herangezogenen Studien, dass ihre Verbreitung deutlich weniger umfangreich ist als oft angenommen und ihr Einfluss weniger stark als medial suggeriert. So zeigen beispielsweise Daten aus dem US-Kontext, dass Desinformation deutlich seltener vorkommt als in der öffentlichen Debatte dargestellt und ebenso selten von Nutzenden geteilt werden (Guess et al., 2023). Ein sehr ähnliches Bild ergibt sich für Deutschland. Die Daten zeigen, dass die Nutzung sogenannter alternativer Nachrichten in Deutschland vor allem auf einen kleinen Teil der Gesellschaft zurückzuführen ist, die regelmäßig auf alternative Nachrichten zugreift (Puschmann et al., 2024). Unter alternativen Nachrichten wird eine Reihe von nicht-mainstream und extremen Medienmarken verstanden, die unter anderem für die Verbreitung von Desinformation bekannt sind.

Die Auswirkungen von Desinformation und anderen systemischen Risiken im Kontext digitaler Plattformen auf die Gesellschaft sind vielfältig und komplex. Die digitale Transformation der politischen Öffentlichkeit führt zu einer konfliktorientierten und negativ-emotionalisierten Verschiebung des öffentlichen Diskurses, da menschliche Wahrnehmungs-Biases und die Aufmerksamkeitsökonomie emotionalisierende, polarisierende Inhalte begünstigen. Dies kann im Zusammenspiel mit weiteren Einflussfaktoren zu einer Verstärkung der gesellschaftlichen Polarisierung beitragen und zentrale Funktionen der politischen Öffentlichkeit beeinträchtigen, indem Desinformationen sowie extremere Ansichten und Narrative verbreitet und Misstrauen gegenüber den gesellschaftlich oder politisch „Anderen“ geschürt werden (Esau et al., 2024). Die gezielte Überflutung des öffentlichen Diskurses mit Provokationen, Tabubrüchen und Desinformationen dient als Strategie des Agenda Hackings, um mediale Aufmerksamkeit zu monopolisieren, politische Gegner zu überfordern und extremistische Narrative zu normalisieren. Zusätzlich führt die Verbreitung von Desinformation als auch die Debatte über potenzielle Desinformationskampagnen zu einem Vertrauensverlust der Bürger:innen in traditionelle Medien, öffentliche Institutionen sowie Funktionsträger:innen der Gesellschaft. Dabei ist Vertrauen ein unverzichtbares Element demokratischer Gesellschaften. Digitale Plattformen haben zwar das Potenzial, politische Prozesse zu bereichern, sie können jedoch auch das Vertrauen in Institutionen und den gesellschaftlichen Zusammenhalt gefährden. Das Zusammenspiel aus

erodierendem Vertrauen in gesellschaftliche Institutionen (Karlsen & Aalberg, 2023) sowie einem von Moral, Emotionen und Gruppenidentitäten geprägten digitalen Diskurs (z.B. Arora et al., 2022) erleichtert es radikalen und extremistischen Akteur:innen, ihre Ansichten und Ideologien zu verbreiten. So können digitale Medien zur Radikalisierung und extremistisch motivierter Gewalt beitragen. Ziel extremistisch motivierter Gewalt werden unter anderem Funktionsträger:innen der Gesellschaft, wie Politiker:innen, Journalist:innen, Wissenschaftlicher:innen und engagierte Bürger:innen auf lokaler, regionaler und nationaler Ebene (z.B. Seeger et al., 2024). Diese Gewalt drückt sich nicht nur in physischen Angriffen aus, sondern im digitalen Raum oft in Form von illegalen Hassnachrichten, Falschbehauptungen sowie der Verbreitung falscher und negativer Narrativer über die betroffene Person. Hierbei sind insbesondere marginalisierte Gruppen überproportional häufig von solchen Anfeindungen betroffen. Das kann für diese zu einer erhöhten psychischen Belastung, Rückzug aus der Öffentlichkeit oder Selbstzensur führen (Seeger et al., 2024). So schwächen Anfeindungen und Falschbehauptungen direkt das Rückgrat einer funktionierenden und lebendigen demokratischen Gesellschaft.

Die sozialwissenschaftliche Messung und Quantifizierung dieser Auswirkungen gestaltet sich angesichts des komplexen Zusammenspiels vieler Faktoren als herausfordernd. Forschende nutzen hierfür unterschiedliche Methoden zur Erfassung der Merkmale von Desinformation, deren Verbreitung sowie der Interaktion der Nutzenden mit diesen Inhalten. Des Weiteren zielen Forschungsstudien auf die Einstellungen von Nutzenden, die tatsächliche Exposition von Desinformation sowie Einflussfaktoren und Wirkungsmechanismen von Desinformation (z.B. Guess et al., 2023; Meyer et al., 2023; Möhring, 2013; Diekmann, 2014). Dieser methodische Werkzeugkasten ermöglicht zwar eine umfassende Abbildung einzelner Aspekte von Desinformation, die Zusammenführung dieser Aspekte sowie die Feststellung von stichhaltigen kausalen Zusammenhängen ist aber äußerst schwierig. Denn die Vielzahl gleichzeitig auftretender Einflussfaktoren und methodische Einschränkungen in der Forschung machen kausale Zuschreibungen problematisch. Daher lässt sich zwar überzeugend argumentieren, dass Desinformationen zu den aufgezählten Herausforderungen beitragen, wie groß die Rolle von Desinformationen gegenüber anderen – potenziell wichtigen – Faktoren ist, lässt sich jedoch schwer bestimmen.

Was als gesichert angesehen werden kann ist, dass die digitale Transformation eine Auswirkung auf den öffentlichen Diskurs und die medial vermittelten politischen Aushandlungsprozesse entfaltet, der unter anderem das Angebot von und die Nachfrage nach radikalen und extremistischen Ansichten und Ideologien befeuert (z.B. Esau et al., 2024). Die Komplexität kommunikativer Meinungsbildungsprozesse auf individueller und gesellschaftlicher Ebene macht es indes unmöglich, Desinformation als alleinige oder auch primäre Ursache für diese Auswirkungen zu identifizieren. Gleichzeitig zeigen repräsentative Daten, dass die meisten Menschen politische Informationen aus vielfältigen Quellen beziehen, von einer Vielzahl politischen Akteur:innen mit unterschiedlichen Standpunkten (Stier et al., 2021). Entsprechend kann davon ausgegangen werden, dass es sich bei Filterblasen und Echokammern eher um randständige und wenig einflussreiche Phänomene in Bezug auf die Polarisierung der Gesellschaft handelt. Vor diesem Hintergrund könnte die in der Gesellschaft wahrgenommene

Verbreitung und Relevanz der Desinformation die wichtigste Einflussgröße sein – und nicht die tatsächlich vorkommende Desinformation. Interessanterweise zeigt sich in der Wahrnehmung von Desinformation, dass Menschen sich selbst als weniger anfällig einschätzen als andere. Dabei sind ältere Menschen, Frauen und Personen mit höherem Bildungsgrad tendenziell stärker besorgt über die Auswirkungen von Desinformation. Die individuelle Wahrnehmung von Desinformation beeinflusst wiederum die Reaktionen und das Verhalten der Betroffenen. Aus diesen Erkenntnissen lässt sich schließen, dass es sich bei Desinformation zwar um ein ernstzunehmendes Problem handelt, was aber in der gesellschaftlichen Debatte möglicherweise überbewertet wird, und dadurch das gesellschaftliche Vertrauen stärker untergraben wird als die tatsächliche Verbreitung falscher Informationen. Ein zentraler Ansatzpunkt für zukünftige Bemühungen sollte daher der Schutz der geteilten gesellschaftlichen Vertrauensbasis sein.

Für die weitere Forschung ergeben sich mehrere vielversprechende Richtungen: Zunächst sollten die langfristigen Auswirkungen von Desinformation auf gesellschaftliche Strukturen und demokratische Prozesse untersucht werden. Dabei wäre es wichtig, robuste Methoden zu entwickeln, um die spezifischen Effekte von Desinformation von anderen Einflussfaktoren zu isolieren. Ein Ansatz wäre es, die in den Election Guidelines und im Code of Conduct on Desinformation enthaltenen Maßnahmen auf langfristige Effektivität zu untersuchen. Weiterhin sollte erforscht werden, wie die Wahrnehmung von Desinformation das Vertrauen in Institutionen und Medien beeinflusst und welche Faktoren die individuelle Resilienz gegenüber Falschinformationen stärken. Schließlich wären Studien zur Wirksamkeit verschiedener Gegenmaßnahmen, wie Medienkompetenzvermittlung oder Fact-Checking-Initiativen, von großem Wert, um effektive Strategien im Umgang mit Desinformation zu entwickeln. Darüber hinaus eröffnet sich wissenschaftlich die Frage, ob der Wahrheitsgehalt einer Äußerung entscheidend ist, wenn der potenzielle Schaden von Inhalten bewertet wird. Auch wahrheitsgemäße Äußerungen können – bei entsprechender (ggf. kontextloser) – Verbreitung zu gesellschaftlich schädlichem Inhalt werden.

Der DSA benennt systemische Risiken in verschiedenen Kategorien: Dazu gehören die Verbreitung rechtswidriger Inhalte, Beeinträchtigungen von Grundrechten, negative Einwirkungen auf gesellschaftliche Debatten und Wahlprozesse sowie Risiken für das körperliche und geistige Wohlbefinden. Diese Kategorisierung bietet jedoch keine abschließende Definition systemischer Risiken. Vielmehr dient sie als Rahmen, um die vielfältigen Gefahrenpotenziale digitaler Dienste zu erfassen. Mit Blick auf die im DSA von VLOPs und VLOSEs geforderte Identifizierung und Minderung systemischer Risiken lässt sich daher die Frage stellen, welche anderen Einflussfaktoren neben Desinformationen einen möglicherweise ertragreichen Zugang bieten. So bildet beispielsweise die im Gutachten dargelegte konfliktorientierte und negativ-emotionalisierte Verschiebung des öffentlichen Diskurses im Kontext der digitalen Aufmerksamkeitsökonomie nicht nur eine wichtige Grundlage für daran anknüpfende Problembereiche wie der digitalen Verbreitung von Desinformationen und anderen Herausforderungen wie politischen Extremismus und gesellschaftlicher Polarisierung. Sondern sie wird oft von Plattform-Architekturen und den implementierten automatisierten Content-Governance-Systemen unmittelbar verstärkt, liegt also - zumindest in Teilen - unmittelbar im

Handlungs- und Verantwortungsbereich der Plattformen. Vielversprechender als die oft mühsame und vor allem verfassungsrechtlich problematische Bekämpfung einzelner Desinformationen könnte ein Fokus auf die strukturellen Bedingungen sein, in denen Desinformationen und andere problematische Inhalte sich verbreiten.

Für eine umfassende Betrachtung der systemischen Risiken im Kontext des DSA ist es daher entscheidend, die Verflechtungen von Akteur:innen wie auch die zentralen Funktionen digitaler Dienste zu berücksichtigen. Plattformen sind nicht nur technische Infrastrukturen, sondern auch soziale Räume, die Einfluss auf die demokratische Meinungsbildung und gesellschaftliche Stabilität haben. Ein angemessenes Verständnis dieser Risiken erfordert also eine ganzheitliche Perspektive auf die Dynamiken der digitalen Kommunikation.

LITERATURVERZEICHNIS

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring Systemic Risk. *Review of Financial Studies*, 30(1), 2–47. <https://doi.org/10.1093/rfs/hhw088>

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, 13(1), 30. <https://doi.org/10.1007/s13278-023-01028-5>

Altay, S. (2023). Who Is Concerned About Misinformation and Why? Evidence From 46 Countries Between 2018 and 2023. *OSF*. <https://doi.org/10.31234/osf.io/fz6wu>

Amnesty International. (2022). *The Social Atrocity: Meta and the Right to Remedy for the Rohingya*.

Arora, S. D., Singh, G. P., Chakraborty, A., & Maity, M. (2022). Polarization and social media: A systematic review and research agenda. *Technological Forecasting and Social Change*, 183, 121942. <https://doi.org/10.1016/j.techfore.2022.121942>

ARD-aktuell. (2024a, 27. Januar). EU verhängt Sanktionen gegen „Voice of Europe“ und weitere russische Medien. *Tagesschau*. <https://www.tagesschau.de/inland/innenpolitik/voice-of-europe-sanktionen-eu-afd-100.html>

ARD-aktuell. (2024b, 26. Januar). Auswärtiges Amt: Desinformationskampagne Russlands aufgedeckt. *Tagesschau*. <https://www.tagesschau.de/inland/desinformation-kampagne-russland-100.html>

Aruguete, N., Calvo, E., & Ventura, T. (2023). News by Popular Demand: Ideological Congruence, Issue Salience, and Media Reputation in News Sharing. *The International Journal of Press/Politics*, 28(3), 558–579. <https://doi.org/10.1177/19401612211057068>

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2024). Online searches to evaluate misinformation can increase its perceived veracity. *Nature*, 625(7995), Article 79–95. <https://doi.org/10.1038/s41586-023-06883-y>

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general*, 149(8), 1608.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>

Baker, P. (2024, November 3). Trump's Wild Claims, Conspiracies and Falsehoods Redefine Presidential Bounds. *The New York Times*. <https://www.nytimes.com/2024/11/03/us/politics/trump-falsehoods-claims-election.html>

Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), 154–175. <https://doi.org/10.1080/21670811.2017.1345645>

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10), 1531–1542.

Barman, D., Guo, Z., & Conlan, O. (2024). The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination. *Machine Learning with Applications*, 16, 100545. <https://doi.org/10.1016/j.mlwa.2024.100545>

Baqir, A., Galeazzi, A., & Zollo, F. (2024). News and misinformation consumption: A temporal comparison across European countries. *PLOS ONE*, 19(5), Article 5. <https://doi.org/10.1371/journal.pone.0302473>

Bayer, J., Katsirea, I., Batura, O., Holznagel, B., Hartmann, S., & Lubianiec, K. (2021). The fight against disinformation and the right to freedom of expression. Report for the European Parliament's Committee on Civil Liberties, Justice and Home Affairs. Universities of Münster, Sheffield, Leuphana, and Ecorys, Poland.

Bayerisches Landesamt für Verfassungsschutz. (2024). „Doppelgänger“: Interne Details zu Russischer Desinformationskampagne. Teil 2 – Vollanalyse.

Beckett, C., & Yaseen, M. (2023). Generating Change. A global survey of what news organisations are doing with AI. The London School of Economics and Political Science. <https://static1.squarespace.com/static/64d60527c01ae7106f2646e9>

Behre, J., Hölig, S., & Möller, J. (2024). Reuters Institute Digital News Report 2024: Ergebnisse für Deutschland. Arbeitspapiere des Hans-Bredow-Instituts. <https://doi.org/10.21241/ssor.94461>

Bennett, W. L., & Pfetsch, B. (2018). Rethinking Political Communication in a Time of Disrupted Public Spheres. *Journal of Communication*, 68(2), 243–253. <https://doi.org/10.1093/joc/jqx017>

Bennett, W. L., & Segerberg, A. (2012). The Logic of Connective Action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, 15(5), 739–768. <https://doi.org/10.1080/1369118X.2012.670661>

Bernhard, L., Schulz, L., Berger, C., & Unzicker, K. (2024). Verunsicherte Öffentlichkeit: Superwahljahr 2024: Sorgen in Deutschland und den USA wegen Desinformationen. Bertelsmann Stiftung. <https://doi.org/10.11586/2024020>

Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*, 14614448241253138. <https://doi.org/10.1177/14614448241253138>

Bitschnaus, S., & Eberspach, K. (2024). KOMMUNALES MONITORING Hass, Hetze und Gewalt gegenüber Amtsträgerinnen und Amtsträgern (KoMo) [Auswertung der Herbstbefragung 2023]. <https://www.staedtetag.de/files/dst/docs/Publikationen/Weitere-Publikationen/2024/KoMo-Herbstbefragung-2023-1.pdf>

Bitton, D. B., Hoffmann, C. P., & Godulla, A. (2024). Deepfakes in the context of AI inequalities: Analysing disparities in knowledge and attitudes. *Information, Communication & Society*, 1–21. <https://doi.org/10.1080/1369118X.2024.2420037>

Bleyer-Simon, K., & Reviglio, U. (2024). Defining Disinformation across EU and VLOP Policies. European Digital Media Observatory. <https://edmo.eu/wp-content/uploads/2024/10/EDMO-Report-%E2%80%93-Defining-Disinformation-across-EU-and-VLOP-Policies.pdf>

Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Dutkiewicz, L., Krack, N., Teysou, D., Severio Nucci, F., Spangenberg, J., Srba, I., Aichroth, P., Cuccovilla, L., & Verdoliva, L. (2024). Generative AI and

Disinformation: Recent Advances, Challenges, and Opportunities. Vera.Ai, AI4Trust, AI4Media, Titan. https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-_White-Paper-v8.pdf

Bouchaud, P., Faddoul, M., & Buse Çetin, R. (2024). No Embargo In Sight. Meta Lets pro-Russian Propaganda Ads Flood the EU. AI Forensics. https://aiforensics.org/uploads/No_Embargo_in_Sight_AI_Forensics_Report_ad7ede416b.pdf

Boulianne, S., & Hoffmann, C. P. (2024). Perceptions and Concerns About Misinformation on Facebook in Canada, France, the US, and the UK. *International Journal of Public Opinion Research*, 36(4), edae048. <https://doi.org/10.1093/ijpor/edae048>

Bovermann, M. (2024). Community Notes auf dem Prüfstand. <https://doi.org/10.59704/d39353743945e3b8>

Blockade in Brasilien: Starlink kündigt Widerstand gegen Sperre von X an. (2024, September 2). *Der Spiegel*. <https://www.spiegel.de/netzwelt/netzpolitik/x-in-brasilien-starlink-kuendigt-widerstand-gegen-sperre-an-a-6c621949-d69b-494e-a6e3-4931b4845942>

Brady, W. J., Crockett, M., & Van Bavel, J. J. (2019). The MAD Model of Moral Contagion: The role of motivation, attention and design in the spread of moralized content online. <https://doi.org/10.31234/osf.io/pz9g6>

Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. J. (2023). Algorithm-mediated social learning in online social networks. *Trends in Cognitive Sciences*, 27(10), 947–960. <https://doi.org/10.1016/j.tics.2023.06.008>

Brandtzaeg, P. B., Følstad, A., & Chaparro Domínguez, M. Á. (2018). How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. *Journalism Practice*, 12(9), 1109–1129. <https://doi.org/10.1080/17512786.2017.1363657>

Breiter, A., & Hepp, A. (2018). Die Komplexität der Datafizierung: Zur Herausforderung, digitale Spuren in ihrem Kontext zu analysieren. *57137*, 4, 27–48. <https://doi.org/10.17174/dcr.v4.2>

Bright, J. (2018). Explaining the Emergence of Political Fragmentation on Social Media: The Role of Ideology and Extremism. *Journal of Computer-Mediated Communication*, 23(1), 17–33. <https://doi.org/10.1093/jcmc/zmx002>

Broggi, E., & De Gregorio, G. (2024). From the code of practice to the code of conduct? Navigating the future challenges of disinformation regulation. *Journal of Media Law*, 16(1), 38–46. <https://doi.org/10.1080/17577632.2024.2362480>

Broughton Micova, S., & Calef, A. (2023). Elements for Effective Systemic Risk Assessment Under the DSA. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4512640>

Broughton Micova, S., & Schnurr, D. (2024). Systemic Risk in Digital Services: Benchmarks for Evaluating the Management of Risks to Electoral Processes. *Center on Regulation in Europe*.

Brüggemann, M., & Meyer, H. (2023). When debates break apart: Discursive polarization as a multi-dimensional divergence emerging in and through communication. *Communication Theory*, 33(2), 132–142. <https://doi.org/10.1093/ct/qtad012>

Bruns, A. (2019). *Are filter bubbles real?* Polity Press.

Bundesnetzagentur (2024). Studie zur Inhaltsmoderation und zur Umsetzung der TCO-VO (EU-VO zur Bekämpfung der Verbreitung terroristischer Online-Inhalte). Goldmedia GmbH im Auftrag der Bundesnetzagentur.

<https://www.bundesnetzagentur.de/DE/Fachthemen/Digitalisierung/Internet/TerrorOnlIn/Studie.pdf>

Butter, M. (2021). Bildung ist der Schlüssel. In Berghof Foundation Operations gGmbH (Hrsg.), Zum kritischen Umgang mit Verschwörungstheorien: Erkenntnisse für die pädagogische Praxis (S. 11–18).

Chadwick, A. (2013). *The Hybrid Media System: Politics and Power*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199759477.001.0001>

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>

Cheruiyot, D. (2024). Comparing Risks to Journalism: Media Criticism in the Digital Hate. *Digital Journalism*, 12(3), 294–313. <https://doi.org/10.1080/21670811.2022.2030243>

Chicago Project on Security & Threats (CPOST). (2022). DEEP, DIVISIVE, DISTURBING AND CONTINUING: NEW SURVEY SHOWS MAINSTREAM COMMUNITY SUPPORT FOR VIOLENCE TO RESTORE TRUMP REMAINS STRONG [Findings from the Fall 2021 CPOST (NORC) American Political Violence Survey]. The University of Chicago. https://d3qi0qp55mx5f5.cloudfront.net/cpost/i/docs/Pape_AmericanInsurrectionistMovement_2022-01-02.pdf?mtime=1641247264

Cinelli, M., Cresci, S., Quattrociocchi, W., Tesconi, M., & Zola, P. (2022). Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems*, 160, 113819. <https://doi.org/10.1016/j.dss.2022.113819>

Citrin, J., & Stoker, L. (2018). Political Trust in a Cynical Age. *Annual Review of Political Science*, 21(Volume 21, 2018), 49–70. <https://doi.org/10.1146/annurev-polisci-050316-092550>

Congress of Local and Regional Authorities. (2022, Oktober 25). Hate speech and fake news—The impact on working conditions of local and regional elected representatives—Rapporteurs Wilma Delissen Van Tongerlo, Netherlands (L, ILDG), Kristoffer Tamsons, Sweden (R, PPE/CCE). Hate speech and fake news: the impact on working conditions of local and regional elected representatives. Congress of Local and Regional Authorities 43rd Session, Budapest. [https://search.coe.int/congress#\[%22CoEIdentifier%22:\[%220900001680a88102%22\],%22sort%22:\[%22CoEValidationDate%20Descending%22\]}](https://search.coe.int/congress#[%22CoEIdentifier%22:[%220900001680a88102%22],%22sort%22:[%22CoEValidationDate%20Descending%22]})

Charta der Grundrechte der Europäischen Union, Pub. L. No. 2012/C 326/02 (2012).

Couldry, N., & Hepp, A. (2023). Die mediale Konstruktion der Wirklichkeit: Eine Theorie der Mediatisierung und Datafizierung. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-37713-7>

Dahlgren, P. M. (2021). A critical review of filter bubbles and a comparison with selective exposure. *Nordicom Review*, 42(1), 15–33. <https://doi.org/10.2478/nor-2021-0002>

De León, E., Makhortykh, M., & Adam, S. (2024). Hyperpartisan, Alternative, and Conspiracy Media Users: An Anti-Establishment Portrait. *Political Communication*, 41(6), 877–902.

<https://doi.org/10.1080/10584609.2024.2325426>

de-Lima-Santos, M.-F., & Ceron, W. (2023). Coordinated Amplification, Coordinated Inauthentic Behaviour, Orchestrated Campaigns: A Systematic Literature Review of Coordinated Inauthentic Content on Online Social Networks. In *Mapping Lies in the Global Media Sphere*. Routledge.

Delhey, J., & Newton, K. (2003). Who trusts?: The origins of social trust in seven societies. *European Societies*, 5(2), 93–137. <https://doi.org/10.1080/1461669032000072256>

Dergacheva, D., Kuznetsova, V., Scharlach, R., & Katzenbach, C. (2023). One Day in Content Moderation: Analyzing 24 h of Social Media Platforms' Content Decisions through the DSA Transparency Database. *Lab Platform Governance, Media, and Technology (PGMT)*. Centre for Media, Communication and Information Research (ZeMKI), University of Bremen. <https://doi.org/10.26092/elib/2707>

DeVerna, M. R., Yan, H. Y., Yang, K.-C., & Menczer, F. (2023). Fact-checking information from large language models can decrease headline discernment (Version 4). arXiv. <https://doi.org/10.48550/ARXIV.2308.10800>

Dierickx, L., & Lindén, C.-G. (2023). Journalism and Fact-Checking Technologies: Understanding User Needs [PDF]. <https://doi.org/10.7275/CPO.1879>

Döring, N., & Mohseni, M. R. (2019). Fail videos and related video comments on YouTube: A case of sexualization of women and gendered hate speech? *Communication Research Reports*, 36(3), 254–264. <https://doi.org/10.1080/08824096.2019.1634533>

Dreyer, S., Stanciu, E., Potthast, K. C., & Schulz, W. (2021). Desinformation. Risiken, Regulierungslücken und adäquate Gegenmaßnahmen. Landesanstalt für Medien NRW.

Duffy, C. (2024). How Elon Musk has turned X into a pro-Trump machine. CNN. <https://www.cnn.com/2024/08/13/tech/elon-musk-donald-trump-x/index.html>

Eder, N. (2023). Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation (SSRN Scholarly Paper No. 4491365). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.4491365>

Efroni, Z. (2021). The Digital Services Act: Risk-based regulation of online platforms. *Internet Policy Review*. <https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606>

Europäische Kommission. (2022, Juni 16). 2022 Strengthened Code of Practice on Disinformation.

Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F. K., Bruns, A., O'Connor-Farfan, K. S., & Lubicz-Zaorski, C. (2024). Destructive polarization in digital communication contexts: A critical review and conceptual framework. *Information, Communication & Society*, 1–22. <https://doi.org/10.1080/1369118X.2024.2413127>

Fazio, L., Rand, D. G., Lewandowsky, S., Susmann, M., Berinsky, A., Guess, A. M., Kendeou, P., Lyons, B., Miller, J., Newman, E., Pennycook, G., & Swire-Thompson, B. (2024). Combating misinformation: A megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headline. <https://doi.org/10.31234/osf.io/uyjha>

Fenton, N., & Barassi, V. (2011). Alternative Media and Social Networking Sites: The Politics of Individuation and Political Participation. *The Communication Review*, 14(3), 179-196.

<https://doi.org/10.1080/10714421.2011.597245>

Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday*, *28*(6), 13185. <https://doi.org/10.5210/fm.v28i6.13185>

Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, *7*(1), 549–569. <https://doi.org/10.1007/s42001-024-00250-1>

Figà Talamanca, G., & Arfini, S. (2022). Through the Newsfeed Glass: Rethinking Filter Bubbles and Echo Chambers. *Philosophy & Technology*, *35*(1), 20. <https://doi.org/10.1007/s13347-021-00494-z>

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. <https://doi.org/10.1177/1461444817724170>

Franck, G. (1998). *Ökonomie der Aufmerksamkeit: Ein Entwurf*.

Gauck, J. (2016, Dezember). Bundespräsident Joachim Gauck beim Festakt anlässlich 60 Jahre Deutscher Presserat am 1. Dezember 2016 in Berlin. <https://www.bundespraesident.de/SharedDocs/Reden/DE/Joachim-Gauck/Reden/2016/12/161201-Festakt-Presserat.html>

Gaudette, T., Scrivens, R., & Venkatesh, V. (2022). The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists. *Terrorism and Political Violence*, *34*(7), 1339–1356. <https://doi.org/10.1080/09546553.2020.1784147>

Geers, M., Swire-Thompson, B., Lorenz-Spreen, P., Herzog, S. M., Kozyreva, A., & Hertwig, R. (2024). The Online Misinformation Engagement Framework. *Current Opinion in Psychology*, *55*, 101739. <https://doi.org/10.1016/j.copsyc.2023.101739>

Gehl, R. W. & Bakardjieva, M. (2016). *Socialbots and Their Friends. Digital Media and the Automation of Sociality*. New York, Abingdon: Routledge.

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*, *101*(7), 2980–3018. <https://doi.org/10.1257/aer.101.7.2980>

Ghai, A., Kumar, P., & Gupta, S. (2024). A deep-learning-based image forgery detection framework for controlling the spread of misinformation. *Information Technology & People*, *37*(2), 966–997. <https://doi.org/10.1108/ITP-10-2020-0699>

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society*, *8*(3), 20563051221117552. <https://doi.org/10.1177/20563051221117552>

González-Bailón, S., & Lelkes, Y. (2023). Do social media undermine social cohesion? A critical review. *Social Issues and Policy Review*, *17*(1), 155–180. <https://doi.org/10.1111/sipr.12091>

Griffin, R. (2023). The Law and Political Economy of Online Visibility. *Technology and Regulation*, 69-79 Pages. <https://doi.org/10.26116/TECHREG.2023.007>

Grolleau, G., Marciano, A., & Mzoughi, N. (2020). The Strategic Use of Scandals. *Kyklos*, *73*(4), 524–542.

<https://doi.org/10.1111/kykl.12249>

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

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature human behaviour*, 4(5), 472–480. <https://doi.org/10.1038/s41562-020-0833-x>

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404. <https://doi.org/10.1126/science.abp9364>

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656), 404–408. <https://doi.org/10.1126/science.add8424>

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10, 178–206. https://doi.org/10.1162/tacl_a_00454

Haas, A., & Unkel, J. (2017). Ranking versus reputation: Perception and effects of search result credibility. *Behaviour & Information Technology*, 36(12), 1285–1298. <https://doi.org/10.1080/0144929X.2017.1381166>

Haidt, J., & Bail, C. (2024). *Social Media and Political Dysfunction: A Collaborative Review*. New York University.

Hameleers, M., Powell, T. E., Van Der Meer, T. G. L. A., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, 37(2), 281–301. <https://doi.org/10.1080/10584609.2019.1674979>

Hameleers, M., Van Der Meer, T. G. L. A., & Brosius, A. (2020). Feeling “disinformed” lowers compliance with COVID-19 guidelines: Evidence from the US, UK, Netherlands, and Germany. *Harvard Kennedy School Misinformation Review*, 1(3). <https://doi.org/10.37016/mr-2020-023>

Hameleers, M., Brosius, A., Marquart, F., Goldberg, A. C., Van Elsas, E., & De Vreese, C. H. (2022). Mistake or Manipulation? Conceptualizing Perceived Mis- and Disinformation among News Consumers in 10 European Countries. *Communication Research*, 49(7), 919–941. <https://doi.org/10.1177/0093650221997719>

Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2022). You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media. *Social Media + Society*, 8(3), 20563051221116346. <https://doi.org/10.1177/20563051221116346>

Hameleers, M., & Marquart, F. (2023). It's Nothing but a Deepfake! The Effects of Misinformation and Deepfake Labels Delegitimizing an Authentic Political Speech. *International Journal of Communication*, 17, 6291–6311.

Hameleers, M., Van Der Meer, T., & Vliegthart, R. (2024a). How persuasive are political cheapfakes disseminated via social media? The effects of out-of-context visual disinformation on message credibility

and issue agreement. *Information, Communication & Society*, 1–18.
<https://doi.org/10.1080/1369118X.2024.2388079>

Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2024b). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior*, 152, 108096. <https://doi.org/10.1016/j.chb.2023.108096>

Hartmann, D., Pohlmann, L., Wang, S. M., & Berendt, B. (2024). A Systematic Review of Echo Chamber Research: Comparative Analysis of Conceptualizations, Operationalizations, and Varying Outcomes. arXiv preprint arXiv:2407.06631

Hashmi, A., Shahzad, S. A., Lin, C.-W., Tsao, Y., & Wang, H.-M. (2024). Unmasking Illusions: Understanding Human Perception of Audiovisual Deepfakes (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2405.04097>

HateAid, Koch, L., Voggenreiter, A. & Steinert, J. I. (2025). Angegriffen & alleingelassen. Wie sich digitale Gewalt auf politisches Engagement auswirkt. Ein Lagebild. <https://hateaid.org/neue-studie-politisch-engagierte-und-digitale-gewalt/>.

Heese, J., Perez Cavazos, G., & Peter, C. D. (2021). When the Local Newspaper Leaves Town: The Effects of Local Newspaper Closures on Corporate Misconduct. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3889039>

Heft, A., & Buehling, K. (2022). Measuring the diffusion of conspiracy theories in digital information ecologies. *Convergence*. <https://doi.org/10.1177/13548565221091809>

Hepp, A. (2018). Von der Mediatisierung zur tiefgreifenden Mediatisierung. In J. Reichertz & R. Bettmann (Hrsg.), *Kommunikation – Medien – Konstruktion* (S. 27–45). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21204-9_2

Hepp, A. (2021). *Auf dem Weg zur digitalen Gesellschaft. Über die tiefgreifende Mediatisierung der sozialen Welt*. Köln: Halem.

Hestermann, T., Hoven, E., & Autenrieth, M. (2021). „Eine Bombe, und alles ist wieder in Ordnung“: Eine Analyse von Hasskommentaren auf den Facebook-Seiten reichweitenstarker deutscher Medien [Application/pdf]. DARIAH-DE. <https://doi.org/10.20375/0000-000E-6320-E>

Hirndorf, D. (2020). Let's talk about politics. Konrad-Adenauer-Stiftung. <https://www.kas.de/de/einzeltitel/-/content/let-s-talk-about-politics>

Hofmann, F., & Raue, B. (2023). *Digital Services Act. Gesetz über digitale Dienste. Nomos*.

Hohlfeld, R. (2020). Wahr oder falsch? Eine empirische Untersuchung zur Wahrnehmung von „Fake News“ und echten Nachrichten in der politischen Kommunikation. In R. Hohlfeld, M. Harnischmacher, E. Heinke, L. Lehner, & M. Sengl (Hrsg.), *Fake News und Desinformation* (S. 179–202). Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783748901334-179>

Hölig, S. (2018). Eine meinungsstarke Minderheit als Stimmungsbarometer?! Über die Persönlichkeitseigenschaften aktiver Twitterer. *Medien & Kommunikationswissenschaft*, 66(2), 140–169. <https://doi.org/10.5771/1615-634X-2018-2-140>

Hooghe, M., & Marien, S. (2013). *A Comparative Analysis of the Relation Between Political Trust and Forms*

of Political Participation in Europe. *European Societies*, 15(1), 131–152. <https://doi.org/10.1080/14616696.2012.692807>

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics*, 25(3), 493–516. <https://doi.org/10.1177/1940161219900126>

Institute for Strategic and Dialogue. (2024, Juli 31). From rumours to riots: How online misinformation fuelled violence in the aftermath of the Southport attack. *Digital Dispatches*. https://www.isdglobal.org/digital_dispatches/from-rumours-to-riots-how-online-misinformation-fuelled-violence-in-the-aftermath-of-the-southport-attack/

Institute for Strategic Dialogue. (2024). Systemisches Risiko für Wahlen? Eine erste Analyse von TikTok-Empfehlungssystemen im Kontext der Landtagswahl Brandenburg.

Iosifidis, P., & Nicoli, N. (2020). The battle to end fake news: A qualitative content analysis of Facebook announcements on how it combats disinformation. *International Communication Gazette*, 82(1), 60–81.

Jack, C. (2017). *Lexicon of Lies: Terms for Problematic Information* (S. 1–20). *Data & Society*.

Jaki, S., & De Smedt, T. (2019). Right-wing German Hate Speech on Twitter: Analysis and Automatic Detection (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1910.07518>

Javed, M., Zhang, Z., Dahri, F. H., & Laghari, A. A. (2024). Real-Time Deepfake Video Detection Using Eye Movement Analysis with a Hybrid Deep Learning Approach. *Electronics*, 13(15), 2947. <https://doi.org/10.3390/electronics13152947>

Jost, P., & Dogruel, L. (2023). Radical Mobilization in Times of Crisis: Use and Effects of Appeals and Populist Communication Features in Telegram Channels. *Social Media + Society*, 9(3), 20563051231186372. <https://doi.org/10.1177/20563051231186372>

Jin, X., Zhang, Z., Gao, B., Gao, S., Zhou, W., Yu, N., & Wang, G. (2023). Assessing the perceived credibility of deepfakes: The impact of system-generated cues and video characteristics. *New Media & Society*, 14614448231199664. <https://doi.org/10.1177/14614448231199664>

Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, 9(3), 20563051231186353. <https://doi.org/10.1177/20563051231186353>

Jungherr, A., Posegga, O., & An, J. (2019). Discursive Power in Contemporary Media Systems: A Comparative Framework. *The International Journal of Press/Politics*, 24(4), 404–425. <https://doi.org/10.1177/1940161219841543>

Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling Politics: How Digital Media Are Shaping Democracy* (1. Aufl.). Cambridge University Press. <https://doi.org/10.1017/9781108297820>

Kaiser, B., & Mayer, J. (2023). It's the Algorithm: A large-scale comparative field study of misinformation interventions. <https://knightcolumbia.org/content/its-the-algorithm-a-large-scale-comparative-field-study-of-misinformation-interventions>

Kaiser, J. & Rauchfleisch, A. (2019). Integrating Concepts of Counterpublics into Generalised Public Sphere Frameworks: Contemporary Transformation in Radical Forms. *Javnost - The Public* 26(3), 241-257.

<https://doi.org/10.1080/13183222.2018.1558676>

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, 23(5), 1301–1326. <https://doi.org/10.1177/1461444820959296>

Karlsen, R., & Aalberg, T. (2023). Social Media and Trust in News: An Experimental Study of the Effect of Facebook on News Story Credibility. *Digital Journalism*, 11(1), 144–160. <https://doi.org/10.1080/21670811.2021.1945938>

Kessler, S. H. (2023). Vorsicht #Desinformation: Die Wirkung von desinformierenden Social Media-Posts auf die Meinungsbildung und Interventionen.

Kettemann, M. C., & Schulz, W. (2023). Platform://Democracy: Perspectives on Platform Power, Public Values and the Potential of Social Media Councils. Verlag Hans-Bredow-Institut. <https://doi.org/10.21241/SSOAR.86524>

Kharel, A., Paranjape, M., & Bera, A. (2023). DF-TransFusion: Multimodal Deepfake Detection via Lip-Audio Cross-Attention and Facial Self-Attention (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2309.06511>

KI-Reifegrad Report. (2024). Bundesverband Digitalpublisher und Zeitungsverleger e.V. (BDZV) und Retresco. https://www.bdzv.de/fileadmin/content/6_Service/6-1_Presse/6-1-2_Pressemitteilungen/2024/PDFs/Retresco_BDZV_KI_Reifegrad_Report_2024_lang.pdf

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>

Knuutila, A., Neudert, L.-M., & Howard, P. N. (2022). Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-97>

Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., Barzilay, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L. K., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., ... Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, 8(6), 1044–1052. <https://doi.org/10.1038/s41562-024-01881-0>

Krempf, S. (2024, Mai 26). Unterstützung für Medien: Google droht mit Stopp der News-Initiative in den USA. heise online. <https://www.heise.de/news/Unterstuetzung-fuer-Medien-Google-droht-mit-Stopp-der-News-Initiative-in-den-USA-9732395.html>

Krotz, F., & Hepp, A. (Hrsg.). (2012). Mediatisierte Welten. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-94332-9>

Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115610141>

Lecheler, S., & Egelhofer, J. L. (2022). Disinformation, Misinformation, and Fake News. In J. Strömbäck, Å. Wikforss, K. Glüer, T. Lindholm, & H. Oscarsson, Knowledge Resistance in High-Choice Information

- Environments (1. Aufl., S. 69–87). Routledge. <https://doi.org/10.4324/9781003111474-4>
- Lee, J., & Shin, S. Y. (2022). Something that They Never Said: Multimodal Disinformation and Source Vividness in Understanding the Power of AI-Enabled Deepfake News. *Media Psychology*, 25(4), 531–546. <https://doi.org/10.1080/15213269.2021.2007489>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 69. <https://doi.org/10.1186/1748-5908-5-69>
- Levi, M., & Stoker, L. (2000). Political Trust and Trustworthiness. *Annual Review of Political Science*, 3(Volume 3, 2000), 475–507. <https://doi.org/10.1146/annurev.polisci.3.1.475>
- Lewis, A., Vu, P., Duch, R. M., & Chowdhury, A. (2023). Deepfake detection with and without content warnings. *Royal Society Open Science*, 10(11), 231214. <https://doi.org/10.1098/rsos.231214>
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, 7(1), 74–101. <https://doi.org/10.1038/s41562-022-01460-1>
- Lübke, S. (2024, Mai 28). Humor in der Klimakrise: Wie Memes den politischen Klimadiskurs beeinflussen | republica. <https://re-publica.com/de/session/humor-der-klimakrise-wie-memes-den-politischen-klimadiskurs-beeinflussen>
- Makhortykh, M., Urman, A., Münch, F. V., Heldt, A., Dreyer, S., & Kettemann, M. C. (2022). Not all who are bots are evil: A cross-platform analysis of automated agent governance. *New Media & Society*, 24(4), 964–981. <https://doi.org/10.1177/14614448221079035>
- Makhortykh, M., & Wijermars, M. (2021). Can Filter Bubbles Protect Information Freedom? Discussions of Algorithmic News Recommenders in Eastern Europe. *Digital Journalism*, 1–25. <https://doi.org/10.1080/21670811.2021.1970601>
- Makridis, C. A. & Thayer, Joel (2023). Data Currency: A Reevaluation of the Consumer Welfare Standard for Digital Markets. <https://dx.doi.org/10.2139/ssrn.4512410>
- Mannocci, L., Mazza, M., Monreale, A., Tesconi, M., & Cresci, S. (2024). Detection and Characterization of Coordinated Online Behavior: A Survey (No. arXiv:2408.01257). arXiv. <https://doi.org/10.48550/arXiv.2408.01257>
- Mansell, R., Durach, F., Kettemann, M., Lenoir, T., Procter, R., Tripathi, G., & Tucker, E. (2025). Information Ecosystems and Troubled Democracy: A Glo9bal Synthesis of the State of Knowledge on News Media, AI and Data Governance. *Observatory on Information and Democracy*. <https://observatory.informationdemocracy.org/reports/>
- Marcks, H. (2023, Juli 24). Jeder wirbt für sich allein? Wie auf Telegram der Aufruhr zum Geschäft wird. *Machine Against the Rage*. <https://machine-vs-rage.bag-gegen-hass.net/jeder-wirbt-fur-sich-allein/>
- Margetts, H., John, P., Hale, S. A., & Yasseri, T. (2017). *Political turbulence: How social media shape collective action* (First paperback printing). Princeton University Press.
- Marini, M. M., & Singer, B. (1988). Causality in the social sciences. *Sociological methodology*, 18, 347–409.

Matthes, J., Corbu, N., Jin, S., Theocharis, Y., Schemer, C., van Aelst, P., Strömbäck, J., Koc-Michalska, K., Esser, F., Aalberg, T., Cardenal, A. S., Castro, L., de Vreese, C., Hopmann, D., Sheaffer, T., Splendore, S., Stanyer, J., Stępińska, A., Štětka, V., & Zoizner, A. (2023). Perceived prevalence of misinformation fuels worries about COVID-19: A cross-country, multi-method investigation. *Information, Communication & Society*, 26(16), 3133–3156. <https://doi.org/10.1080/1369118X.2022.2146983>

Mauk, M., & Grömping, M. (2024). Online Disinformation Predicts Inaccurate Beliefs About Election Fairness Among Both Winners and Losers. *Comparative Political Studies*, 57(6), Article 6. <https://doi.org/10.1177/00104140231193008>

Mayring, P. (2004). 5.12 Qualitative Content Analysis. *A companion to qualitative research*, 266–269.

Mayring, P., & Fenzl, T. (2019). *Qualitative Inhaltsanalyse*. Springer.

McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). Misinformation exploits outrage to spread online. *Science*, 386(6725), 991–996. <https://doi.org/10.1126/science.adl2829>

Mercier, H. (2021). How Good Are We At Evaluating Communicated Information? *Royal Institute of Philosophy Supplements*, 89, 257–272. <https://doi.org/10.1017/S1358246121000096>

Meta. (2024). 2024 European Parliament Post-Elections Report. Digital Services Act—Elections Guidelines.

Mercier, H. (2021). How Good Are We At Evaluating Communicated Information? *Royal Institute of Philosophy Supplements*, 89, 257–272. <https://doi.org/10.1017/S1358246121000096>

Meyer, H., Peach, A. K., Guenther, L., Kedar, H. E., & Brüggemann, M. (2023). Between calls for action and narratives of denial: Climate change attention structures on Twitter. *Media and Communication*, 11(1), 278–292.

Mirza, S., Begum, L., Niu, L., Pardo, S., Abouzied, A., Papotti, P., & Pöpper, C. (2023). Tactics, Threats & Targets: Modeling Disinformation and its Mitigation. *Proceedings 2023 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium, San Diego, CA, USA. <https://doi.org/10.14722/ndss.2023.23657>

Möhring, W. (2013). *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Springer VS.

Mondak, J. (1995). *Nothing to Read: Newspapers and Elections in a Social Experiment*. University of Michigan Press. <https://doi.org/10.3998/mpub.10442>

Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dutse, I., Khan, S., & Parkinson, S. (2023). A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats. *IEEE Access*, 11, 144497–144529. <https://doi.org/10.1109/ACCESS.2023.3344653>

Nadimpalli, A. V., & Rattani, A. (2024). Social Media Authentication and Combating Deepfakes using Semi-fragile Invisible Image Watermarking. <https://doi.org/10.48550/ARXIV.2410.01906>

Nakamura, D. (2016). Media critic Obama is worried that ‘balkanized’ media is feeding partisanship. *The Washington Post*. https://www.washingtonpost.com/politics/media-critic-obama-is-worried-that-balkanized-media-are-feeding-partisanship/2016/03/27/8c72b408-f1e3-11e5-89c3-a647fcce95e0_story.html

- Nenno, S., & Lorenz-Spreen, P. (2025). Musk, X und die AfD: Profitiert Alice Weidel? Digital Society Blog. <https://doi.org/10.5281/zenodo.14749544>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2021). Reuters Institute Digital News Report 2021. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021>
- Nielsen, R. K. (2019). Economic Contexts of Journalism. In K. Wahl-Jorgensen & T. Hanitzsch (Hrsg.), *The handbook of journalism studies* (2nd edition). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781315167497>
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., & Dimmery, D. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620, 137–144. <https://doi.org/10.1038/s41586-023-06297-w>
- Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-024>
- Pacelli, V., Panetta, I. C., & Povia, M. M. (2024). Systemic Risk and Network Science: A Bibliometric and Systematic Review. In V. Pacelli (Hrsg.), *Systemic Risk and Complex Networks in Modern Financial Systems* (S. 21–42). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64916-5_2
- Papendick, M., Rees, Y. P. M., Wäschle, F., & Zick, A. (2020). Hass und Angriffe auf Medienschaffende—Eine Studie zur Wahrnehmung von und Erfahrungen mit Angriffen auf Journalist*innen [Forschungsbericht IKG]. Universität Bielefeld - Institut für interdisziplinäre Konflikt- und Gewaltforschung (IKG). https://pub.uni-bielefeld.de/download/2943243/2943245/Studie_Hass_und_Angriffe_auf_Medienschaffende.pdf
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Park, S., Fisher, C., Flew, T., & Dulleck, U. (2020). Global Mistrust in News: The Impact of Social Media on Trust. *International Journal on Media Management*, 22(2), 83–96. <https://doi.org/10.1080/14241277.2020.1799794>
- Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., Aluvala, S., & Vimal, V. (2023). Deepfake Generation and Detection: Case Study and Challenges. *IEEE Access*, 11, 143296–143323. <https://doi.org/10.1109/ACCESS.2023.3342107>
- Pedrazzi, S., & Oehmer, F. (2023). Governance of misinformation by public service bots: The potential of automated social networking site accounts operated by public service media for correcting misinformation. Nordicom, University of Gothenburg. <https://doi.org/10.48335/9789188855756-10>
- Puschmann, C., Stier, S., Zerrer, P., & Rauxloh, H. (2024). Politicized and paranoid? Assessing attitudinal predictors of alternative news consumption. *Journal of Broadcasting & Electronic Media*, 68(4), 489–518.
- Quandt, T., Klapproth, J., & Frischlich, L. (2022). Dark social media participation and well-being. *Current Opinion in Psychology*, 45, 101284. <https://doi.org/10.1016/j.copsyc.2021.11.004>
- Ranaivoson, H., Afilipoaie, A., & Domazetovikj, N. (2022). Media pluralism in the EU: A prospective look at the European Media Freedom Act. Policy Brief No. 64. Vrije Universiteit Brussel, Studies in Media Innovation Technology Research Group.

Rau, J. P. (2023). Digitale Turbulenzen: Aufbau einer demokratischen Gesellschaft in Zeiten des digitalen Umbruchs. In: Institut für Demokratie und Zivilgesellschaft (Hg.). Wissen schafft Demokratie. Schwerpunkt Netzkulturen und Plattformpolitiken, Band 14. Jena, 16–29.

Rau, J., Kero, S., Hofmann, V., Dinar, C., & Heldt, A. P. (2022): Rechtsextreme Online-Kommunikation in Krisenzeiten. Herausforderungen und Interventionsmöglichkeiten aus Sicht der Rechtsextremismus- und Platform-Governance-Forschung. Hamburg: Hans-Bredow-Institut, März 2022 (Arbeitspapiere des Hans-Bredow-Instituts | FGZ Resultate Nr. 62). <https://doi.org/10.21241/ssoar.78072>

Rau, J. P., & Stier, S. (2019). Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien? Zeitschrift für Vergleichende Politikwissenschaft, 13(3), 399–417. <https://doi.org/10.1007/s12286-019-00429-1>

Reiss, M., Stöwing, E., Knor, E., Merten, L., & Möller, J. (2025, im Erscheinen). Generative künstliche Intelligenz zur Informationsnavigation. Verlag Hans-Bredow-Institut (Arbeitspapiere des Hans-Bredow-Instituts).

Reporters Without Borders. (2020). 2020 World Press Freedom Index: Entering a decisive decade for journalism, exacerbated by coronavirus. Reporters Without Borders. <https://rsf.org/en/2020-world-press-freedom-index-entering-decisive-decade-journalism-exacerbated-coronavirus>

Righetti, N., Giglietto, F., Kulichkina, A., Marino, G., & Terenzi, M. (2022). Politische Werbung und koordiniertes Verhalten in sozialen Medien im Vorfeld der Bundestagswahl 2021.

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering Misinformation: Evidence, Knowledge Gaps, and Implications of Current Interventions. European Psychologist, 28(3), 189–205. <https://doi.org/10.1027/1016-9040/a000492>

Röper, H. (2022). Zeitungsmarkt 2022: Weniger Wettbewerb bei steigender Konzentration. Media Perspektiven, 6, 295–318.

Ross Arguedas, A., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). Echo chambers, filter bubbles, and polarisation: A literature review. Reuters Institute for the Study of Journalism. <https://doi.org/10.60625/RISJ-ETXJ-7K60>

Ross, A. S., & Rivers, D. J. (2019). Internet Memes, Media Frames, and the Conflicting Logics of Climate Change Discourse. Environmental Communication, 13(7), 975–994. <https://doi.org/10.1080/17524032.2018.1560347>

Roth-Isigkeit, D. (2024). Der risikobasierte Ansatz als Paradigma des Digitalverwaltungsrechts. Multimedia und Recht (MMR).

Rovira Kaltwasser, C., & Mudde, C. (2017). Populism: A Very Short Introduction. Oxford University Press. <https://doi.org/10.1093/actrade/9780190234874.001.0001>

Ruscheimer, H., Quintais, J. P., Nenadić, I., De Gregorio, G., & Eder, N. (2024). Brave New World. <https://doi.org/10.59704/46b8611eb2d96a84>

Schroeder, R. (2018). Social Theory after the Internet: Media, Technology, and Globalization. UCL Press. <https://doi.org/10.2307/j.ctt20krxdr>

Schulz, W., & Ollig, C. (2023). Hybrid Speech Governance: New Approaches to govern Social Media Platforms under the European Digital Services Act? JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law, 14(4), Article 4. <https://www.jipitec.eu/jipitec/article/view/22>

Schulze, A., Brand, F., Leschzyk, D. K., Beuthner, M., Biegert, A., Bomnüter, U., Boy, B., Bucher, H.-J., Frau, R., Hubig, M., Löffelholz, M., Mayer, J., Pliquet, C., Radechovsky, J., Schleicher, K., & Ulbrich, K. (2023). Optimierung der Risiko- und Krisenkommunikation von Regierungen, Behörden und Organisationen der Gesundheitssicherung – Herausforderungen in lang anhaltenden Krisen am Beispiel der COVID-19-Pandemie. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 66(8), 930–939. <https://doi.org/10.1007/s00103-023-03708-1>

Schüz, B., & Jones, C. (2024). Falsch- und Desinformation in sozialen Medien: Ansätze zur Minimierung von Risiken in digitaler Kommunikation über Gesundheit. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 67(3), 300–307. <https://doi.org/10.1007/s00103-024-03836-2>

Schwaiger, L. (2022). Gegen die Öffentlichkeit: Alternative Nachrichtenmedien im deutschsprachigen Raum. transcript. <https://doi.org/10.14361/9783839461211>

Schweiger, W. (2017). Nachrichtenjournalismus, alternative und soziale Medien. Der (des)informierte Bürger im Netz. Springer. https://doi.org/10.1007/978-3-658-16058-6_2

Shin, S. Y., & Lee, J. (2022). The Effect of Deepfake Video on News Credibility and Corrective Influence of Cost-Based Knowledge about Deepfakes. Digital Journalism, 10(3), 412–432. <https://doi.org/10.1080/21670811.2022.2026797>

Siegrist, M., & Cvetkovich, G. (2000). Perception of Hazards: The Role of Social Trust and Knowledge. Risk Analysis, 20(5), 713–720. <https://doi.org/10.1111/0272-4332.205064>

Sievers, P., Mayer, C., & Kahlmann, A. (2020). Standortanalyse der Zustellung für Tageszeitungen in Deutschland. SCHICKLER Unternehmensberatung GmbH im Auftrag des Bundesverbands Digitalpublisher und Zeitungsverleger (BDZV) e.V., Berlin.

Silva, T. C., Da Silva, M. A., & Tabak, B. M. (2017). Systemic risk in financial systems: A feedback approach. Journal of Economic Behavior & Organization, 144, 97–120. <https://doi.org/10.1016/j.jebo.2017.09.013>

Sippy, T., Enock, F., Bright, J., & Margetts, H. Z. (2024). Behind the Deepfake: 8% Create; 90% Concerned. Surveying public exposure to and perceptions of deepfakes in the UK (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2407.05529>

Slechten, L., Courtois, C., Coenen, L., & Zaman, B. (2021). Adapting the Selective Exposure Perspective to Algorithmically Governed Platforms: The Case of Google Search. Communication Research, 00936502211012154. <https://doi.org/10.1177/00936502211012154>

Snyder, J. M., & Strömberg, D. (2010). Press Coverage and Political Accountability. Journal of Political Economy, 118(2), 355–408. <https://doi.org/10.1086/652903>

Sønderskov, K. M., & Dinesen, P. T. (2016). Trusting the State, Trusting Each Other? The Effect of Institutional Trust on Social Trust. Political Behavior, 38(1), 179–202. <https://doi.org/10.1007/s11109-015-9322-8>

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic

Vigilance. *Mind & Language*, 25(4), 359–393. <https://doi.org/10.1111/j.1468-0017.2010.01394.x>

Stark, B., Magin, M., & Jürgens, P. (2021). Maßlos überschätzt. Ein Überblick über theoretische Annahmen und empirische Befunde zu Filterblasen und Echokammern. In M. Eisenegger, M. Prinzing, P. Ettinger, & R. Blum (Hrsg.), *Digitaler Strukturwandel der Öffentlichkeit: Historische Verortung, Modelle und Konsequenzen* (S. 303–321). Springer Fachmedien. https://doi.org/10.1007/978-3-658-32133-8_17

Steppat, D., Castro Herrero, L., & Esser, F. (2022). Selective exposure in different political information environments – How media fragmentation and polarization shape congruent news use. *European Journal of Communication*, 37(1), 82–102. <https://doi.org/10.1177/02673231211012141>

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—Revised. *Annual review of clinical psychology*, 11, 71–98.

Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., Posch, L., Maier, J., Rothmund, T., & Staab, S. (2018). Systematically monitoring social media: The case of the German federal election 2017 (GESIS Papers). GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.56149>

Stier, S., Kirkizh, N., Froio, C., & Schroeder, R. (2020). Populist Attitudes and Selective Exposure to Online News: A Cross-Country Analysis Combining Web Tracking and Surveys. *The International Journal of Press/Politics*, 25(3), 426–446. <https://doi.org/10.1177/1940161220907018>

Stier, S., Mangold, F., Scharnow, M., & Breuer, J. (2021). Post post-broadcast democracy? News exposure in the age of online intermediaries. *American Political Science Review*, 1–7. <https://doi.org/10.1017/s0003055421001222>

Strübing, J., Hirschauer, S., Ayaß, R., Krähnke, U., & Scheffer, T. (2018). Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie*, 47(2), 83–100. <https://doi.org/10.1515/zfsoz-2018-1006>

Sun, H. (2023). *Regulating Algorithmic Disinformation*.

Tajrian, M., Rahman, A., Kabir, M. A., & Islam, Md. R. (2023). A Review of Methodologies for Fake News Analysis. *IEEE Access*, 11, 73879–73893. <https://doi.org/10.1109/ACCESS.2023.3294989>

Tandoc, E. C. (2021). Fake news. In *The Routledge companion to media disinformation and populism* (S. 110–117). Routledge.

Tsang, S. J., & Rojas, H. (2020). Opinion Leaders, Perceived Media Hostility and Political Participation. *Communication Studies*, 71(5), 753–767. <https://doi.org/10.1080/10510974.2020.1791203>

Thorson, K., & Wells, C. (2016). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age: Curated Flows. *Communication Theory*, 26(3), 309–328. <https://doi.org/10.1111/comt.12087>

TikTok. (2024, Juli 9). How we kept our community safe during the 2024 European Parliament elections. Newsroom | TikTok. <https://newsroom.tiktok.com/en-eu/how-we-kept-our-community-safe-during-the-2024-european-parliament-elections>

Transparency International. (2020). *Corruption Perceptions Index 2020*. <https://www.transparency.org/en/cpi/2020>

Urman, A., & Makhortykh, M. (2023). You are how (and where) you search? Comparative analysis of web search behavior using web tracking data. *Journal of Computational Social Science*, 6(2), 741–756. <https://doi.org/10.1007/s42001-023-00208-9>

Van Aelst, P., & Walgrave, S. (2016). Information and Arena: The Dual Function of the News Media for Political Elites: Information and Arena. *Journal of Communication*, 66(3), 496–518. <https://doi.org/10.1111/jcom.12229>

Van Der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How Can Psychological Science Help Counter the Spread of Fake News? *The Spanish Journal of Psychology*, 24, e25. <https://doi.org/10.1017/SJP.2021.23>

Van Raemdonck, N., & Meyer, T. (2024). Why disinformation is here to stay. A socio-technical analysis of disinformation as a hybrid threat. In L. Lonardo (Hrsg.), *Addressing Hybrid Threats* (S. 57–83). Edward Elgar Publishing. <https://doi.org/10.4337/9781802207408.00009>

Vegetti, F., & Mancosu, M. (2022). Perceived Exposure and Concern for Misinformation in Different Political Contexts: Evidence From 27 European Countries. *American Behavioral Scientist*, 00027642221118255. <https://doi.org/10.1177/00027642221118255>

Verhulst, S. G. (2023). Steering Responsible AI: A Case for Algorithmic Pluralism (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2311.12010>

Verordnung (EU) 2023/2854 des Europäischen Parlaments und des Rates vom 13. Dezember 2023 über harmonisierte Vorschriften für einen fairen Datenzugang und eine faire Datennutzung sowie zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie (EU) 2020/1828.

Verordnung (EU) 2024/1083 des Europäischen Parlaments und des Rates vom 11. April 2024 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt und zur Änderung der Richtlinie 2010/13/EU.

Verordnung (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. Mai 2022 über europäische Daten-Governance und zur Änderung der Verordnung (EU) 2018/1724 (2022).

Verordnung (EU) 2022/1925 des Europäischen Parlaments und des Rates vom 14. September 2022 über bestreitbare und faire Märkte im digitalen Sektor und zur Änderung der Richtlinien (EU) 2019/1937 und (EU) 2020/1828 (2022).

Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (2022).

Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (2024).

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>

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, 69, 102199. <https://doi.org/10.1016/j.electstud.2020.102199>

Wagner, M. (2024). Affective polarization in Europe. *European Political Science Review*, 16(3), 378–392. <https://doi.org/10.1017/S1755773923000383>

Wagner, A., & Schwarzenegger, C. (2020). A populism of lulz. The proliferation of humor, satire, and memes as populist communication in digital culture. *Perspectives on Populism and the Media. Avenues for Research*, 7. <https://pub.uni-bielefeld.de/record/2954932>

Wannagat, Simon: Musks Megafon als Parteispende?: Influencer-Beiträge und Werbemaßnahmen Dritter im Parteiengesetz, *VerfBlog*, 2025/1/23, <https://verfassungsblog.de/musks-megafon-als-parteispende/>, DOI: 10.59704/83ce51e73c8a51af.

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making* (No. 27; S. 1–107). Council of Europe.

Washington Post Fact Checker. (2021, Januar 20). Analysis | Tracking all of President Trump's false or misleading claims. *Washington Post*. <https://www.washingtonpost.com/graphics/politics/trump-claims-database/>

Watolla, A.-K. M. (2024). *Kompetenzen kommunikativen Handelns im Kontext mediatisierter Digitalität: Theoriegeleitete Entwicklung eines Kompetenzstrukturmodells für die Lehrkräftebildung* (1. Aufl., Bd. 68). transcript Verlag. <https://doi.org/10.14361/9783839473115>

Webster, J. G. (2014). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. The MIT Press. <https://doi.org/10.7551/mitpress/9892.001.0001>

Weikmann, T., Greber, H., & Nikolaou, A. (2024). After Deception: How Falling for a Deepfake Affects the Way We See, Hear, and Experience Media. *The International Journal of Press/Politics*, 19401612241233539. <https://doi.org/10.1177/19401612241233539>

Wells, C., Shah, D. V., Pevehouse, J. C., Yang, J., Pelled, A., Boehm, F., Lukito, J., Ghosh, S., & Schmidt, J. L. (2016). How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning. *Political Communication*, 33(4), 669–676. <https://doi.org/10.1080/10584609.2016.1224416>

Williams, E. M., & Carley, K. M. (2023). Search engine manipulation to spread pro-Kremlin propaganda. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-112>

World Economic Forum. (2024). *The Global Risks Report 2024* (No. 19).

World Economic Forum. (2025). *The Global Risks Report 2025* (No. 20).

Wuttig, A., & Kessler, S. (2024). *Kurzstudie 2024. Demokratie und Bildung*. <https://static.iu.de/studies/demokratie-und-bildung.pdf>

Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is Inevitable: An Innate Limitation of Large Language Models (Version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2401.11817>

Yamamoto, M., Hmielowski, J. D., Beam, M. A., & Hutchens, M. J. (2018). Skepticism as a political orientation factor: A moderated mediation model of online opinion expression. *Journal of Information Technology & Politics*, 15, 178–192. <https://doi.org/10.1080/19331681.2018.1460287>

Yan, P., & Schroeder, R. (2024). *Drifting Away from the Mainstream: Media Attention and the Politics of*

Hyperpartisan News Websites. *Weizenbaum Journal of the Digital Society*, 4(2). <https://doi.org/10.34669/WI.WJDS/4.2.2>

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, 38(1–2), 98–139. <https://doi.org/10.1080/10584609.2020.1785067>

Zimmermann, F., & Kohring, M. (2020). Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election. *Political Communication*, 37(2), 215–237. <https://doi.org/10.1080/10584609.2019.1686095>

Zmerli, S. (2023). Political Trust. In F. Maggino (Hrsg.), *Encyclopedia of Quality of Life and Well-Being Research* (S. 5278–5281). Springer International Publishing. https://doi.org/10.1007/978-3-031-17299-1_2202