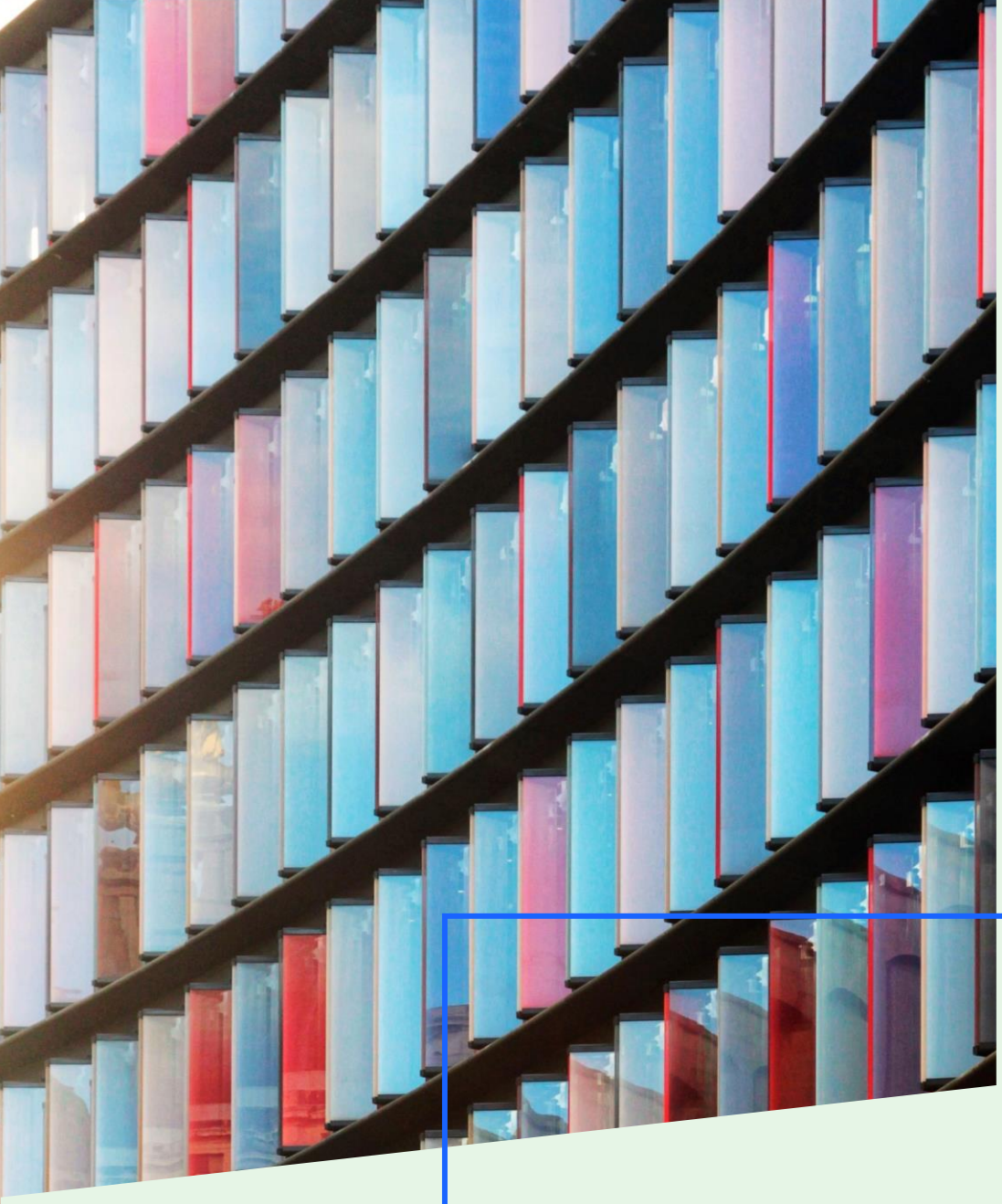


bitkom

Policy Brief

Maßnahmen gegen Desinformation & Deepfakes
im Superwahljahr 2024

Auf einen Blick

Im Vorfeld der US-Wahlen wurde versucht, durch manipulierte Computerstimmen, die Präsident Joe Biden imitierten, Menschen vom Wählen abzuhalten. In der Türkei teilte Präsident Erdoğan ein Video, in dem sein Hauptkonkurrent bei den Parlaments- und Präsidentenwahlen 2023 mit dem Vertreter einer als terroristische Vereinigung eingestuften Organisation gezeigt wurde. Das Video war aus separaten Teilen zusammengeschnitten worden, um fälschlicherweise terroristische Unterstützung des Konkurrenten anzudeuten. Im Netz zirkulierte im vergangenen Jahr ein Video des Bundeskanzlers, Olaf Scholz, der vermeintlich einen Verbotsantrag für die AfD stellte. Das Video wurde als Deepfake identifiziert und gesperrt.

Diese plakativen Beispiele aus der jüngsten Vergangenheit zeigen, wie schnell und effektiv Desinformation Vertrauen zersetzt, Themen schnell eskalieren und negative gesellschaftliche Auswirkungen haben können. Desinformation verfälscht die Basis zur freien Meinungsbildung. Obwohl Desinformation hauptsächlich vor dem Hintergrund politischer Einflussnahme diskutiert wird, ist auch die Wirtschaft vermehrt Ziel von Desinformationskampagnen geworden.

- Die Mitglieder des Bitkom sind sich ihrer Verantwortung bewusst und treten illegalen und schädlichen Inhalten durch vielseitige Maßnahmen entschieden entgegen.
- Die Rolle wirtschaftlicher Akteure ist dabei zweigeteilt: Einerseits nehmen sie eine wichtige Schnittstellenfunktion bei der Moderation von Inhalten ein. Andererseits bietet die Digitalwirtschaft Innovationen und neue technische Lösungen, insbesondere bei der Deepfakebekämpfung.
- Mit Blick auf die Bekämpfung von Desinformation und Deepfakes handelt es sich um eine gesamtgesellschaftliche Aufgabe, bei der Politik, Wirtschaft und nicht staatliche Akteure aus dem Bereich der Forschung, der Zivilgesellschaft und der Medien effektiv zusammenarbeiten sollten. Wie konkrete Kampagnen gegen Desinformation unter Beteiligung von Tech-Unternehmen aussehen können, zeigt die Initiative der Bayern-Allianz gegen Desinformation.
- Das Ziel dieses Papiers ist, die Debatte rund um Desinformation und Deepfakes wissenschaftlich und praxisnah einzuordnen und hiermit einen Grundstein für die weitere Entwicklung neuer Lösungsansätze zu liefern.
- Der Policy Brief stellt hierzu einige Maßnahmen gegen Desinformation und Deepfakes vor, die Unternehmen und Mitglieder des Bitkom bereits heute ergreifen.

67%

Laut einer Studie des Bitkom zum Nachrichtenkonsum im Internet gaben 67 % der Befragten an, Desinformation insbesondere zu Themen, wie Migration, dem Ukraine-Krieg und dem Israel-Gaza Konflikt wahrgenommen zu haben.

Inhaltsverzeichnis

I. Desinformation als Mittel zur Einflussnahme

1. Definition und Abgrenzung

Im wissenschaftlichen, gesellschaftlichen und politischen Diskurs hat sich das Verständnis rund um Desinformation generisch weiterentwickelt. Bekannt wurde die Terminologie im Zusammenhang mit dem populär genutzten Begriff »Fake News« im Kontext der politischen Einflussnahme auf den US-Wahlkampf im Jahr 2016 durch russische Troll-Farmen oder die Einflussnahme auf das BREXIT-Referendum 2020. Zuletzt definierte der Global Risks Report 2024 des Weltwirtschaftsforums Desinformation als das größte Risiko der nächsten 2 Jahre.[1]

Desinformation meint »nachweislich falsche oder irreführende Informationen, die zum wirtschaftlichen Vorteil oder zur absichtlichen Täuschung und Schädigung der Öffentlichkeit erstellt, präsentiert und verbreitet wird«.[2] Um bloße Information von Desinformation zu unterscheiden, kommt es demnach auf drei abstrakte Merkmale an: (1) Die Intention bzw. Manipulationsabsicht, (2) die Motivation und (3) und das damit verbundene Ziel.

Intention: In diesem Zusammenhang ist Desinformation grundsätzlich von Misinformation zu unterscheiden. Hierbei handelt es sich um falsche oder missverständliche Information, die ohne Intention in den Umlauf gebracht wird und ohne dass sich die inverkehrbringende Person über die Inkorrektheit der Information bewusst ist.[3] Einer der Hauptgründe für das Teilen von Misinformation Fehlinformation ist der sogenannte »Confirmation Bias«. Dieser beschreibt die grundsätzliche Neigung von Menschen, Informationen so zu ermitteln, auszuwählen und zu interpretieren, dass diese das eigene Meinungsbild bestätigen. Daneben gelten sozialer Druck, Gruppenzugehörigkeit und soziale Identität zu den Hauptgründen für das Teilen und Verbreiten von Desinformation. Das Vorhandensein einer bewussten Täuschungsabsicht erlaubt es demnach, Desinformation von faktisch nicht korrekten Inhalten abzugrenzen. Dies betrifft etwa Kunstformen, wie Satire, Parodie, Comedy oder Ironie.

Motivation: Aufgrund der systematischen Eigenschaft der Desinformationsverbreitung fällt auch die Motivation und das damit verbundene Ziel bei genauerer Betrachtung ins Gewicht. Im politischen Kontext verfolgt Desinformation grundsätzlich das Ziel der Destabilisierung politischer Systeme. Es wird daher vom europäischen Gesetzgeber auch zunehmend als »systemisches Risiko« verstanden. Im sicherheitspolitischen

[1] World Economic Forum: Global Risks Report 2024, 19. Edition, Januar 2024, Global Risks Report 2024 | World Economic Forum | World Economic Forum (weforum.org)
[2] OECD - Observatory of Public Sector Innovation: Misinformation and Disinformation: An international effort using behavioral science to tackle the spread of misinformation, Public Governance Policy Paper, 19.10.2022, b7709d4f-en.pdf (oecd-ilibrary.org).
[3] Altay, Sacha et. al: A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field, Harvard Kennedy School Misinformation Review, Juli 2023, altay_survey_expert_views_misinfo_20230727.pdf (harvard.edu).

Bereich wird Desinformation als hybride Bedrohung eingeordnet und Informationssicherheit in diesem Zusammenhang als schützenswertes Gut bewertet. Desinformation schwächt die Informations- und Meinungsbildungskultur insgesamt, da sie die Legitimität und das Vertrauen in öffentliche Stellen und deren Kommunikation infrage stellt. Daher spricht man im außen- und sicherheitspolitischen Kontext auch von »Foreign Information Manipulation« (FIMI). Desinformation kann in diesem Zusammenhang von »Propaganda« unterschieden werden, welches sich konzeptionell auf eine spezifische Gruppe und einen territorialen Kontext begrenzt[4].

Grundsätzlich ist zu beachten, dass Desinformation auch wirtschaftlich motiviert sein kann und hierdurch destabilisierende Effekte auf die Gesellschaft insgesamt hat. Sie kann darüber hinaus von einzelnen Personen ausgehen oder einem Netzwerk an Personen und Organisationen. Diese können zudem staatlichen und auch nicht staatlichen Akteuren zugeordnet werden.

Ziel: Das Ziel von Desinformation besteht nach der gängigen Definition darin, einen systematischen Schaden herbeizuführen. Im politischen Kontext steht deshalb auch die Koordinierung von Gegenmaßnahmen mit Fokus auf großangelegte Desinformationskampagnen, mit denen ein erheblicher Schaden für die Gesellschaft und öffentliche Güter einhergeht, im Vordergrund. Dies betrifft unter anderem den Bereich der Krisenkommunikation oder den Schutz der öffentlichen Gesundheit. Mit dem Ziel, einen erheblichen Schaden herbeizuführen, wie etwa bei einer gezielten Wahlmanipulation, ist demnach ein anderer Grad beziehungsweise eine andere Schwere verbunden, die sich von individuellen Fällen von Desinformationen klar unterscheidet.

Der Kern von Desinformation als Konzept ist mithin nicht neu, sondern so alt wie die Menschheit und das Zusammenleben innerhalb politischer Systeme. Die Entstehung neuer Verbreitungswege von Information (in Echtzeit) birgt enorme Potenziale, aber macht die Abgrenzung von Misinformation zu Desinformation zu einer enormen konzeptionellen und praktischen Herausforderung, die immer eine Einzelfallentscheidung bedarf. Hinzu kommen neue Herausforderungen durch den Missbrauch von generativer KI zur Erstellung von Deepfakes, welche die Herausforderungen mit Blick auf den Vertrauensverlust in Informationen, sowie Bild- und Audiomaterial zusätzlich amplifizieren.

Die Etablierung eines einheitlichen Grundverständnisses über Desinformation und die Abgrenzung zu anderen Begriffen ist für die erfolgreiche Erkennung und Operationalisierung von zentraler Bedeutung. Ferner hilft eine einheitliche Begrifflichkeit in der Kommunikation und dabei gezielt Aufmerksamkeit in der Öffentlichkeit für die Dynamiken und negativen Auswirkungen von Desinformation herzustellen. Dies ist daneben auch eine wichtige Voraussetzung für die Einordnung von effektiven Maßnahmen zur Bekämpfung von Desinformation.

[4] Stanley, Jason: How Propaganda Works, Princeton University Press, 2014, https://www.jstor.org/stable/44508992.

X Fakten über Desinformation

- Desinformation bezieht sich auf nachweislich falsche oder irreführende Information, die gezielt in Umlauf gebracht wird.
- Sie kann objektiv wahre Information beinhalten, aber in einem anderen Kontext dargestellt werden, sodass sie unwahre Narrative erzeugt.
- Per se sind absichtliche Falschaussagen nicht rechtswidrig, werden aber als besonders schädlich betrachtet und können eine hybride Bedrohung für demokratische Gesellschaften und Wirtschaftssysteme darstellen.
- Die Auswirkungen von Desinformation sind nur schwer messbar und im Verhältnis zu anderen hybriden Bedrohungen, wie Cyberangriffen, kaum erkennbar.
- Desinformation kann politisch oder wirtschaftlich motiviert sein.
- Die Verbreitung kann analog oder über digitale Medien und Plattformen stattfinden.
- Desinformation geht sowohl von staatlichen als auch nicht staatlichen Akteuren, Individuen, Gruppen und ganzen Netzwerken aus.
- Sie ist kein neues Phänomen, jedoch in einer zunehmend komplexer strukturierten Informationsgesellschaft schwer eingrenzbar.
- Deepfakes können als technische Methode genutzt werden, um Desinformation zu erzeugen. Die Herausforderungen rund um Deepfakes steigern den Bedarf nach Informationsintegrität und Vertrauen in digitale Inhalte zusätzlich.

2. Desinformation im Superwahljahr 2024

2.1 Desinformation im politischen Kontext

Desinformation wird mit dem Ziel der Destabilisierung politischer Systeme eingesetzt, indem vor allem die Legitimität und das Vertrauen in staatliche Akteure und ihre öffentliche Kommunikation untergraben wird. Weiterhin sind die Medienlandschaft und zivilgesellschaftliche Akteure vermehrt Ziel von Desinformationskampagnen geworden. Durch Desinformation profitieren Akteure von einer zunehmenden Polarisierung der Gesellschaft, der Radikalisierung oder der politischen Beeinflussung einzelner.[5]

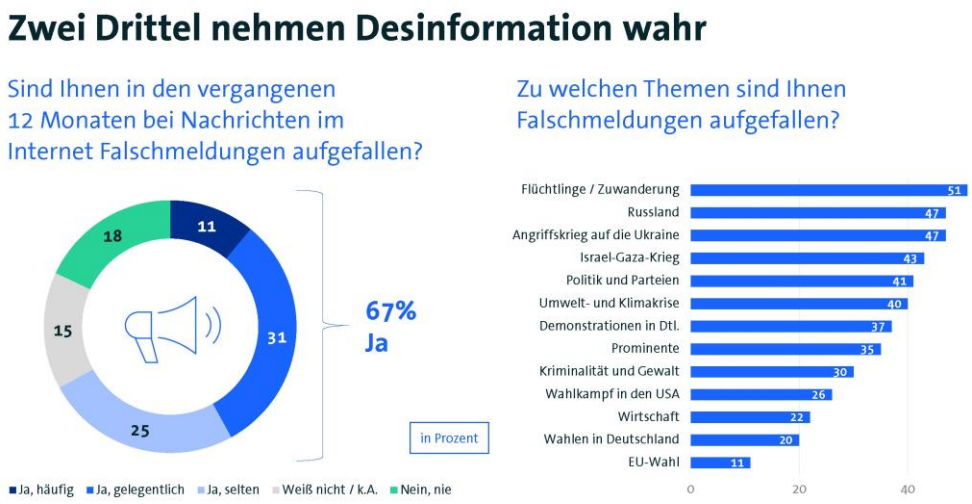
Laut einem Bericht des EU-Sonderausschusses zur Einflussnahme aus dem Ausland stellt Desinformation zudem ein wesentliches Mittel der Wahlbeeinflussung dar.[6]

[5] Statista: Welche Folgen von Missinformation/Desinformation sehen Sie weltweit/in Deutschland?, 2021, Desinformation - gesellschaftliche Auswirkungen 2021 | Statista.
[6] Entschließung des Europäischen Parlaments vom 9. März 2022 zur Einflussnahme aus dem Ausland auf alle demokratischen Prozesse in der Europäischen Union, einschließlich Desinformation (2020/2268(INI)), 2022.

Abgeordnete des Europäischen Parlaments und des Deutschen Bundestages erwarten verstärkte Einmischung und Informationsmanipulation, insbesondere Fälle von bezahlter Desinformation bei Wahlen.

Unabhängig davon nimmt die Menge der Fälle an Informationsmanipulation aus dem Ausland im außen- und sicherheitspolitischen Kontext laut einem Bericht des European External Action Service (EEAS) zu.[7] Hiermit versuchen staatliche Akteure, zum Beispiel aus Russland, China oder dem Iran, auf das öffentliche Meinungsbild einzuwirken und politische Entscheidungen (zum Beispiel mit Blick auf den Krieg in der Ukraine) indirekt zu beeinflussen.

Laut einer repräsentativen Studie des Bitkom zum Nachrichtenkonsum im Internet gaben 67 Prozent der Befragten an, Desinformation insbesondere zu Themen wie Migration, dem Krieg gegen die Ukraine und dem Israel-Gaza-Konflikt wahrgenommen zu haben.[8]

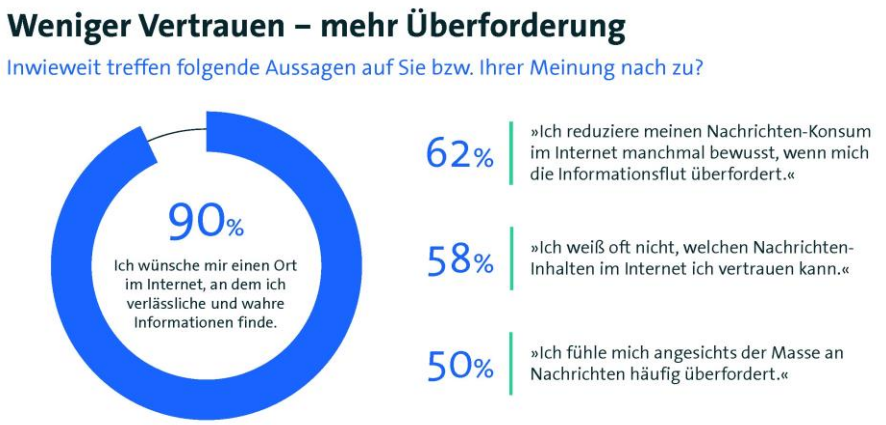


Basis: *Personen, die Online-Nachrichten konsumieren (n=898) | rechts: Personen, denen Fake News aufgefallen sind (n=641) | rechts: Mehrfachnennungen möglich | Quelle: Bitkom Research 2024*

Auch die Medienlandschaft, die Garant für eine freie, demokratische, objektive und pluralistische Nachrichten- und Informationsverbreitung ist, wird durch Desinformation in ihrer Vertrauenswürdigkeit bedroht. Laut der Bitkom-Studie zum Nachrichtenkonsum im Internet informieren sich etwa 90 Prozent der Befragten online.

[7] Strategic Communications, Task Forces and Information Analysis (STRAT.2) Data Team: 1st EEAS Report on Foreign Information Manipulation and Interference Threats, 2023, EEAS-DataTeam-ThreatReport-February2023-02.pdf (europa.eu).
[8] Bitkom Research, Pressekonferenz zur Studie – Nachrichtenkonsum im Internet, 22. Mai 2024, Nachrichtenkonsum im Internet - 2024 (bitkom.org).

Über die Überschriften geht es jedoch oft nicht hinaus. Die Tatsache, dass 88 Prozent der Befragten finden, Online-Medien tendierten dazu, oft reißerische Überschriften zu nutzen, um Klicks zu bekommen, unterstreicht jedoch auch ein allgemeines Misstrauen gegenüber Online-Informationen.



Basis: *Personen, die Online-Nachrichten konsumieren (n=898) | Prozentwerte für »Trifft voll und ganz zu« und »Trifft eher zu« | Quelle: Bitkom Research 2024*

58 Prozent der Befragten gaben an, dass sie oft nicht wissen, welchen Nachrichten sie im Internet vertrauen können. 50 Prozent der Befragten fühlen sich angesichts der Masse an Nachrichten überfordert.

Basis: *Internetnutzerinnen und -nutzer ab 16 Jahren (n=1.002) | Prozentwerte für »Trifft voll und ganz zu« und »Trifft eher zu« | Quelle: Bitkom Research 2024*

Obwohl unsere Gesellschaft also immer mehr zur Informationsgesellschaft insbesondere im Digitalbereich geworden ist, ist auch das Vertrauen in das Informationsmedium »Internet« und die Verlässlichkeit der dort anzutreffenden Informationsangebote gesunken.

2.2 Desinformation in der Wirtschaft

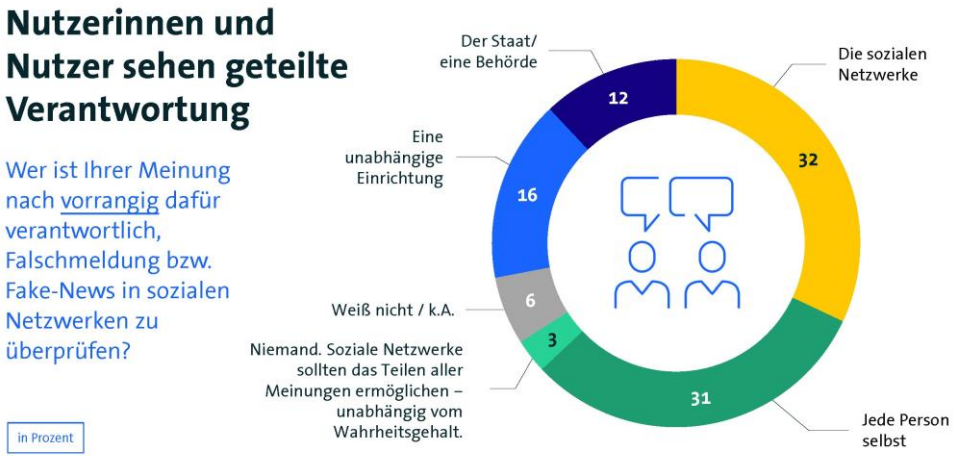
Obwohl Desinformation hauptsächlich vor dem Hintergrund politischer Einflussnahme diskutiert wird, ist auch die Wirtschaft vermehrt Ziel von Desinformationskampagnen geworden. Dies umfasst auf übergeordneter Ebene die Beeinflussung des Verständnisses über wirtschafts- und strukturpolitische Fragen, die einen wesentlichen Bestandteil der Stabilität einer Gesellschaft ausmachen. Darüber hinaus kann Desinformation konkret negative Folgen in Wirtschaftssektoren auslösen, zum Beispiel durch die Manipulation von Aktienkursen oder -märkten.

Desinformation ist ein Einfallstor für Cybersicherheitsrisiken. Ein Beispiel sind Hack-Leak-Kampagnen, bei denen Desinformation als Phishing-Instrument verwendet wird, um sensible Daten von Unternehmen abzugreifen. Diese werden im Nachgang manipuliert und weiterverarbeitet oder als Instrument für Internetkriminalität (z. B. Erpressung oder Betrug) verwendet. Zusätzlich ergeben sich neue Herausforderungen durch generative Künstliche Intelligenz, da der Missbrauch der Technologie die Qualität und Quantität von Social-Engineering-Angriffen und Cyberangriffen steigert (z. B. sprachbasierte Betrugsmethoden durch Voice-Phishing). Eine weitere Herausforderung in diesem Bereich stellt die Fälschung von Onlineidentitäten und der Dokumentenbetrug sowie die Störung von Geschäftsprozessen dar. Sofern dies kritische Unternehmen betrifft, können solche Eingriffe weitreichende Folgen haben. Aber auch kleine und mittelständische Unternehmen haben vermehrt mit solchen Herausforderungen im Geschäftsalltag zu kämpfen.

3. Maßnahmen gegen Desinformation

Die Herausforderung bei der Bekämpfung von Desinformation liegt darin, dass sie in vielen Fällen nicht illegal bzw. rechtswidrig ist. Sie kann deshalb auch nicht erfolgreich allein durch Regulierung adressiert werden. Es besteht jedoch schon heute ein vielfältiges Repertoire an Maßnahmen gegen Desinformation, die in Zusammenwirkung einen effektiven Beitrag zur Bekämpfung liefern. In diesem Zusammenhang betrachten wir die aktuellen Herausforderungen rund um Deepfakes gesondert, da hier unterschiedliche Voraussetzungen der rechtlichen Handhabung vorliegen und sich andere Gegenmaßnahmen ableiten lassen.

Mit Blick auf die Bekämpfung von Desinformation und Deepfakes handelt es sich um eine gesamtgesellschaftliche Aufgabe. Mit Blick auf politische und wirtschaftliche Desinformation ist es essenziell, dass Politik, Wirtschaft, nicht staatliche Akteure aus dem Bereich der Forschung, der Zivilgesellschaft und Medien effektiv zusammenarbeiten. Maßnahmen müssen verstärkt Individuen dabei unterstützen und befähigen, mit den neuen Herausforderungen besser umzugehen. Dies entspricht auch den Ergebnissen der Bitkom-Studie zum Nachrichtenkonsum im Internet, die verdeutlicht, dass die Befragten sich selbst auch in der Eigenverantwortung sehen.



Basis: *Internetnutzerinnen und -nutzer ab 16 Jahren (n=1.002) | Quelle: Bitkom Research 2024*

Im Folgenden wollen wir Maßnahmen gegen Desinformation beleuchten und die Rolle der Digitalwirtschaft in diesem Zusammenhang einordnen. Um eine Grundlage hierfür zu schaffen, geht das folgende Kapitel zunächst auf die rechtliche Handhabung und Aspekte der Regulierung nach derzeitigem Stand ein.

3.1 Rechtliche Handhabung und Regulierung

Die Regulierung von Desinformation steht in einem Spannungsverhältnis mit der Meinungsäußerungs- und Pressefreiheit nach Art. 5 GG. bzw. der Artikel 11 EU-Charta und 10 EMRK. Meinungen sind persönliche Ansichten. Sie sind geschützt von der Meinungsfreiheit, solange die Rechte anderer nicht verletzt werden. Meinungen können also auch rechtswidrig sein. Für die gesellschaftliche Meinungsbildung sind verlässliche Fakten zentral. Diese Fakten müssen klar, nachweisbar und überprüfbar sein. Wichtig ist, dass Meinungen nicht als Fakten dargestellt werden. Desinformation tut genau das, indem vorsätzlich falsche, irreführende oder aus dem Kontext gerissene Behauptungen verbreitet werden. Der (regulatorische) Umgang mit Desinformation ist demnach stets im Lichte dieses Grundrechts sowie der Grundrechte aufseiten der Betroffenen (z. B. Menschenwürde, Persönlichkeitsrecht, Vertraulichkeit von Kommunikation und Privatleben, Schutz der Familie, Recht auf Reputation eines Unternehmens usw.) zu sehen. Desinformation kann deshalb von einer Meinung konzeptionell und rechtlich abgegrenzt werden, die auch missinformiert und damit schädlich, aber nicht rechtswidrig ist. Inhalte sind dann rechtswidrig, wenn die verbreitete Information für sich selbst einen eigenen Tatbestand der Rechtswidrigkeit erfüllt. Unter dieses rechtliche Rahmenwerk fallen in erster Linie Inhalte, die strafrechtlich relevant sind (z. B. Ehrverletzungen, terroristische Inhalte, Kinderpornografie). Ebenfalls infrage kommen persönlichkeits- bzw. datenschutzrechtliche Rechtsbehelfe oder solche aus dem Immaterialgüter- oder Lauterkeitsrecht.

An diese Abgrenzung knüpft auch der Digital Services Act (DSA) an, indem er zwischen rechtswidrigen und schädlichen Inhalten unterscheidet. Rechtswidrige Inhalte umfassen alle Inhalte, Produkte und Dienstleistungen, »die als solche oder durch ihre

Bezugnahme auf eine Tätigkeit, einschließlich des Verkaufs von Produkten oder der Erbringung von Dienstleistungen, nicht im Einklang mit dem Unionsrecht oder dem Recht eines Mitgliedstaats stehen, ungeachtet des genauen Gegenstands oder der Art der betreffenden Rechtsvorschriften« (Art. 3 Bst. h DSA). Nach dem Willen des Gesetzgebers soll der Begriff des illegalen Inhaltes weit gefasst werden. Dabei werden von sich aus illegale Inhalte ebenso erfasst wie Informationen, die sich auf eine illegale Tätigkeit beziehen. Desinformation ist also auch nach dem DSA nicht als solche rechtswidrig.

Desinformation als Gegenstand des Digital Services Acts

Der europäische Gesetzgeber versteht Desinformation auch als »systemisches Risiko« und grenzt dieses insofern klar von rechtswidrigen Inhalten ab. Nutzerinnen und Nutzer, die irreführende Absichten verfolgen, verwenden auch Onlinedienste zum gezielten Streuen von nachweislich falschen oder irreführenden Informationen. Daher fallen insbesondere sogenannten Intermediären und Inhalteanbietern eine besondere Rolle bei der Bekämpfung von Desinformation zu.

Demnach werden vor allem sehr große Online-Plattformen und Suchmaschinen verpflichtet, eine Risikobewertung vorzunehmen und systemische Risiken zu analysieren (Art. 34 (1) DSA). Hierdurch wird zunächst angestrebt, die erforderliche Wissensbasis über Herausforderungen rund um Desinformation herzustellen, auch um die Effektivität von Gegenmaßnahmen besser einschätzen zu können. Neben rechtswidrigen Inhalten fällt hierunter auch die intendierte Manipulation der Services durch eine unauthentische Nutzung, die sich negativ auf schützenswerte öffentliche Güter (Gesundheit, öffentliche Sicherheit, demokratische Wahlprozesse) sowie fundamentale Menschenrechte auswirkt.

Obwohl Desinformation hier nicht konkret als systemisches Risiko benannt wird, lässt es sich unter den Begriff eindeutig subsumieren. Der DSA greift demnach Aspekte der gängigen Definition auf, nach der vorwiegend solche Inhalte, die bewusst und gezielt in den Umlauf gebracht werden, um öffentlichen Schaden herbeizuführen, als Desinformation verstanden werden. Als Beispiel wird auf die negativen Folgen für die öffentliche Gesundheit, öffentliche Sicherheit, oder die Beeinflussung demokratische Wahlprozesse verwiesen.

Sehr große Online-Plattformen und Suchmaschinen sind nach den Vorgaben des DSA auch dazu verpflichtet, Maßnahmen zur Risikominimierung hinsichtlich schädlicher Inhalte, zum Beispiel durch die Durchsetzung der Geschäftsbedingungen und Moderation von Inhalten (Art. 35 (1) b) DSA), vorzunehmen. Darüber hinaus gewährleisten Anbieter Transparenz hinsichtlich der Funktionsweise ihrer Dienste, sodass Nutzerinnen und Nutzer Inhalte besser einordnen können.

Das Instrument der Selbstverpflichtungen

Die Wirkung des DSA wird durch den bereits seit 2018 bestehenden Verhaltenskodex, der seit 2022 weitere Unterzeichner und Maßnahmen gegen Desinformation umfasst, flankiert. Unterzeichner des Kodexes haben sich selbst zur Eindämmung von Desinformation verpflichtet, darunter einige Mitgliedsunternehmen des Bitkom. Sie erfassen Manipulationen und Auswirkungen von Desinformation, bieten mehr

Möglichkeiten für Nutzerinnen und Nutzer, falsche oder irreführende Inhalte zu melden, gewährleisten Transparenz mit Blick auf politische Werbung, engagieren sich im Bereich Faktencheck und bieten einen Datenzugriff für Forschende.[9]

Diese Initiativen werden durch vielfältige Maßnahmen aus Politik, Wirtschaft, Medien und Zivilgesellschaft unterstützt. Um eine bessere Einordnung der Leistung, die die Digitalwirtschaft unternimmt, vorzunehmen, geht das folgende Kapitel auch auf die Rolle anderer Stakeholder ein. Die Digitalwirtschaft und insbesondere die Mitglieder des Bitkom sind sich ihrer Verantwortung hinsichtlich der Bekämpfung von Desinformation bewusst und gehen dies auch in Kooperation mit Partnern aus der Wirtschaft, den Medien und der Zivilgesellschaft aktiv an.

3.2 Governance gegen Desinformation

Da Desinformation nur bedingt durch Regulierung begegnet werden kann, nehmen technische und nicht technische Maßnahmen eine zentrale Rolle bei der Bekämpfung von Desinformation ein. Diese wirken entlang der Entstehungskette von Desinformation – im Bereich der Prävention, bei der Verbreitung, bis hin zur strategischen Gegenkommunikation:

Pre-Bunking: Desinformation wird vorbeugend durch die Sensibilisierung der Öffentlichkeit, Information- und Bildungskampagnen sowie Angebote zur Aufklärung bekämpft. Dies beinhaltet auch die Vermittlung von Hintergrundwissen zu Verbreitungsmechanismen von Desinformation, welches zur kritischen Auseinandersetzung mit Medien beitragen soll. In diesem Zusammenhang ist der langfristige Aufbau von Medienkompetenz entscheidend. Hierdurch wird die Gesellschaft mit Blick auf die kritische Auseinandersetzung mit Information und Nachrichten insgesamt gestärkt.

Fact-Checking: Faktenprüfung ist ein Prozess, der Informationen auf ihre Korrektheit und Wahrheitsgehalt hin überprüft. Dies wird häufig von Journalisten und Journalistinnen, Nachrichtenredaktionen und politischen Analysten durchgeführt. Fact-Checking unterstützt dabei, objektiv korrekte und falsche Information zu unterscheiden und zu korrigieren. Dies ist oft ein langwieriger Prozess, der viele Ressourcen bündelt.

Debunking: Hieran anknüpfend bezieht sich Debunking (entlarven) auf die Richtigstellung, nachdem Desinformation bereits als nachweislich falsche Information identifiziert wurde. Das übergeordnete Ziel besteht darin, die Auswirkungen potenziell schädlicher Desinformation zu minimieren, nachdem Desinformation bereits in den Umlauf geraten ist. Die Effektivität einer Richtigstellung von Inhalten ist jedoch nach aktuellem Stand in der wissenschaftlichen Auseinandersetzung fraglich. Es gibt darüber hinaus Studien, die sogar belegen, dass Debunking einen selbstverstärkenden

[9] EU-Verhaltenskodex zur Bekämpfung von Desinformation, EU-Verhaltenskodex zur Bekämpfung von Desinformation | Europäische Kommission (europa.eu).

Effekt mit Blick auf Desinformation hat, da die Aufmerksamkeit auf als »inkorrekt« deklarierte Inhalte nur noch verstärkt wird.[10]

Strategische Kommunikation: Strategische Kommunikation bezieht sich auf den zielgerichteten Einsatz von Kommunikation durch eine Organisation zur Erfüllung ihres Auftrags.[11] Dieses Konzept geht über die bloße Richtigstellung von Informationen hinaus und zielt darauf ab, ein qualitativ hochwertiges Informationsangebot zu schaffen und Informationslücken zu füllen.

Maßnahmen gegen Desinformation sind oft nicht technischer Natur, da ein automatisiertes Filtern nach nachweislich falschem Content aufgrund der hohen rechtlichen Anforderungen eine praktische Herausforderung darstellt. Dies erfordert, wie bereits erläutert, immer eine Einzelfallabwägung. Unabhängig davon können technische Mittel und Hilfs-Tools aber unterstützen, insbesondere bei der Deepfake-Bekämpfung hinsichtlich der Verifizierung der Echtheit und Herkunft von Inhalten. Aufgrund der Besonderheit der Herausforderungen rund um Deepfakes widmet sich ein gesondertes Kapitel diesem Thema.

Rolle der Politik

Aufgrund der Flut an Desinformation bei gleichzeitiger Zunahme der Masse an Informationen, die im Online-Raum zirkuliert, fokussieren sich staatliche und öffentliche Akteure auf besonders schädliche und großangelegte Desinformationskampagnen. Diese lassen sich leichter identifizieren, da sie oft bestimmten Mustern folgen, von bestimmten Akteuren ausgehen und großflächig angelegt sind. Großangelegte Überwachungs- und Koordinierungsmaßnahmen werden in diesem Zusammenhang von einem Multi-Level-Governance-Netzwerk vorgenommen: Auf europäischer Ebene insbesondere vom European External Action Service (EEAS) und der East StratCom Task Force, welche sich bereits seit 2015 insbesondere mit russischer Desinformation beschäftigt. Die Einordnung von Desinformation als hybride Bedrohung wurde 2016 das erste Mal in einem »Gemeinsamen Rahmen für die Abwehr hybrider Bedrohungen« festgehalten. Hierdurch wurde auch die Stärkung der Resilienz der Europäischen Union und der Ausbau an Partnerschaften zur Bekämpfung hybrider Bedrohung erstmals in den Blick genommen. Eine wichtige Errungenschaft ist auch die Einrichtung von Schnellwarnsystemen (Rapid Alert System der EU und der G7 Rapid Response Mechanism), um den Austausch von Erkenntnissen über Desinformationskampagnen zu erleichtern. Das RAS basiert auf Informationen aus offenen Quellen und greift auf Erkenntnisse aus der Wissenschaft, von Faktenprüfern, Online-Plattformen und internationalen Partnern zurück.[12]

[10] Arcos, Rubén et al.: Responses to digital disinformation as part of hybrid threats: a systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking, Open Research Europe, 2022, Responses to digital disinformation as part ... | Open Research Europe (europa.eu).
[11] Bolte, Neville: Strategic Communications and Disinformation in the Early 21st Century, Robert Schuman Centre for Advanced Studies Research Paper, 2021, https://ssrn.com/abstract=4149452.https://ssrn.com/abstract=4149452https://ssrn.com/abstract=4149452.
[12] European External Action Service: Factsheet: Rapid Alert System, 05.03.2019, Factsheet: Rapid Alert System | EEAS (europa.eu).

Auch auf nationaler Ebene wird mittlerweile eine verstärkte Kooperation beteiligter Bundesbehörden, wie dem Auswärtigen Amt, dem Bundesinnenministerium und dem Bundespresseamt etabliert. Die ressortübergreifende Taskforce gegen Desinformation analysiert insbesondere Desinformation im Kontext des Ukrainekriegs und fokussiert sich in der Arbeit auf eine Stärkung der faktenbasierten Kommunikation, insbesondere von öffentlichen Stellen.[13] Daneben werden weitreichende nationale und multinationale Kooperationseinheiten, wie das European Digital Media Observatory (EDMO) oder das German-Austrian Digital Media Observatory (GADMO) aufgebaut. Es handelt sich dabei um Netzwerke, die darauf abzielen, Desinformation zu bekämpfen und ihre Auswirkungen auf Gesellschaft und Demokratien sowohl auf nationaler als auch auf europäischer Ebene zu analysieren und operativ richtigzustellen.

Daneben spielt die Erhöhung der gesellschaftlichen Resilienz gegen Bedrohungen aus dem Informationsraum eine immer größere Rolle. Die Vermittlung von Medienkompetenz durch Weiterbildungsangebote über verschiedene Demografien hinweg ist eine öffentliche Aufgabe, die weiter ausgebaut werden muss, um nachhaltige Mechanismen gegen Desinformation zu etablieren.

Rolle der Zivilgesellschaft und Wissenschaft

Die Zivilgesellschaft und Wissenschaft nehmen eine zentrale Rolle bei der Bekämpfung von Desinformation ein. Dies betrifft einerseits die Weiterentwicklung der Erkenntnisse über Desinformation und seine Auswirkungen als auch die Bewertung der Effektivität von Gegenmaßnahmen. Darüber hinaus sind Zivilgesellschaft und Wissenschaft essenzieller Bestandteil der Gegenmoderation von Desinformation. Sie genießen oft größeres Vertrauen, sodass auch von ihnen ausgehende oder unter ihrer Beteiligung ergriffene Gegenmaßnahmen wie Fact-Checking, Debunking und strategische Kommunikation besser aufgenommen werden und damit ein Informations-Gegengewicht zur Desinformation darstellen.

Rolle der Medien

Die Sicherung von Medienpluralismus und Medienvielfalt sind essenzielle Bestandteile der Desinformationsbekämpfung, da ein diverses Angebot an qualitativ hochwertiger und objektiver Information die Wahrscheinlichkeit, dass Desinformation wahrgenommen wird, reduziert. Darüber hinaus erfüllen öffentlich-rechtliche und private Medien eine wichtige Aufgabe, da sie Vertrauen in die Qualität und Richtigkeit von Informationen fördern.

Rolle der Wirtschaft

Die digitale Wirtschaft spielt eine entscheidende Rolle bei der Bekämpfung von rechtswidrigen und schädlichen Inhalten, insbesondere von Desinformation oder Hatespeech. Durch verschiedene Maßnahmen stellen sich Online-Plattformen und

[13] Bundesministerium des Inneren und für Heimat, Maßnahmen der Bundesregierung gegen Desinformation, BMI - Alle Schwerpunkte - Maßnahmen der Bundesregierung gegen Desinformation.

Suchmaschinen, aber auch Software und IT-Services, Telekommunikations- oder Internetdienste gegen schädliche Inhalte.

Die Mitglieder des Bitkom sind sich ihrer Verantwortung bewusst und treten neben illegalen auch schädlichen Inhalten durch vielseitige Maßnahmen entschieden entgegen.

Die Rolle wirtschaftlicher Akteure ist dabei zweigeteilt: Auf der einen Seite nehmen sie eine wichtige Schnittstellenfunktion bei der aktiven Bekämpfung, zum Beispiel mit Blick auf die Moderation von Inhalten, wahr. Dies betrifft neben großen Akteuren auch zunehmend kleinere Plattformen, die die Interaktion ihrer Nutzerinnen und Nutzer managen und damit dem Missbrauch ihrer Services vorbeugen wollen. Auf der anderen Seite bietet die Digitalwirtschaft Innovationen und neue technische Lösungen hinsichtlich der Bekämpfung von Desinformation entlang ihrer Entstehungskette. Dies ist insbesondere bei der Deepfake-Bekämpfung von essenzieller Bedeutung. Auf diesen Aspekt geht daher ein gesondertes Kapitel ein. In diesem Kapitel wollen wir zunächst einige Maßnahmen gegen Desinformation, die Unternehmen und Mitglieder des Bitkom bereits heute ergreifen, vorstellen. Hierbei handelt es sich um eine Übersicht, aber nicht abschließende Auflistung an Maßnahmen, die Unternehmen aus dem Bereich der Mitgliedschaft des Bitkom ergreifen:

- **Entfernung von illegalen und schädlichen Inhalten und Fake-Accounts:** Online-Plattformen und Suchmaschinen kommen ihren Verpflichtungen aus gesetzlichen Vorgaben, den eigenen Community-Richtlinien sowie Nutzungsbedingungen ihrer Dienste nach, indem sie aktiv illegale und schädliche Information entfernen. Darüber hinaus können Accounts gesperrt werden. Zentrale Praxis ist jedoch die Moderation von Inhalten. Hierbei gehen Onlinedienste aktiv gegen illegale sowie schädliche Inhalte vor. Unternehmen wollen ein vertrauenswürdiges und sicheres Online-Umfeld schaffen, um eine positive Nutzererfahrung zu gewährleisten. Dies betrifft sowohl das Verhältnis unter den Nutzerinnen und Nutzern als auch die Interaktion im B2C- und B2B-Bereich.

- **Individuelle Anpassung der Dienste**: Über Einstellungen können Nutzerinnen und NutzerInhalte bis zu einem gewissen Grad selbst moderieren. Privatsphäre-Einstellungen bieten insbesondere beim Schutz von Kindern und Jugendlichen einen vorbeugenden Schutzmechanismus. Darüber hinaus können bestimmte Funktionen und Sicherheits-Toolkits dabei helfen, Inhalte mit bestimmten Hashtags oder Kommentaren herauszufiltern, die Nutzerinnen und Nutzer nicht sehen möchten. Diese Funktionen helfen, unerwünschte Inhalte auszublenden.

- **Labeling:** Darüber hinaus labeln viele Anbieter nicht verifizierte Inhalte und schränken die Auffindbarkeit dieser Inhalte durch einen Ausschluss aus Empfehlungsalgorithmen ein. Darüber hinaus kann auch die Erkennbarkeit von schädlicher Desinformation erhöht werden, die oft bestimmten Mustern folgt und insbesondere zu stark polarisierenden Themen (wie dem Ukrainekrieg, dem Gaza-Konflikt) verbreitet wird.

- **Suchintervention:** Bestimmte Inhalte-Kategorien, die auf illegale oder schädliche Inhalte abstellen, werden von den Suchkriterien der Plattformen ausgeschlossen, sodass sie Nutzerinnen und Nutzern erst gar nicht angezeigt werden (z. B. zu Terrorismus und Gewalt).

- **Anzeige von verlässlichen Quellen:** Immer mehr Anbieter machen Inhalte von verlässlichen Quellen besonders sichtbar oder stellen qualitativ hochwertige Informationen zum Thema z. B. durch Links oder Nutzeranmerkungen bereit und schaffen damit ein Gegengewicht zur Desinformation.
- **Einschränkung politischer Werbung:** Einige Akteure sehen Einschränkungen für politische Werbung vor: Politische Inhalte können z.B. auf TikTok nicht beworben werden. Zudem gibt es eingeschränkte Targeting-Optionen oder das Verbot von Micro-Targeting.
- **Transparenz bei politischer Werbung:** Auf einigen Plattformen müssen wiederum Werbetreibende, die Anzeigen schalten einen Autorisierungsprozess durchlaufen und einen Disclaimer einfügen, der die Finanzierung der Anzeige offenlegt. Diese Anzeigen werden dann in öffentlich zugänglichen Anzeigenbibliothek gespeichert. Auf dieses Vorgehen greifen zum Beispiel Meta und Google zurück.
- **Transparenzberichte:** Anbieter von Online-Plattformen veröffentlichen Transparenzberichte und legen dar, wie sie ihre Community-Richtlinien und Nutzungsbedingungen durchsetzen. Hierzu zählt auch die Nachvollziehbarkeit von Auskunftsersuchen im Rahmen der Strafverfolgung, behördliche Aufforderungen zur Entfernung von rechtswidrigen Inhalten und Aufforderungen zur Entfernung von Inhalten im Zusammenhang mit dem Recht an geistigem Eigentum. Darüber hinaus haben Onlineplattformen die Herausforderungen rund um die strategische Einflussnahme im politischen Kontext erkannt und arbeiten gemeinsam mit Behörden auf nationaler und europäischer Ebene an der Identifizierung von Desinformationskampagnen. In diesem Zusammenhang veröffentlichen Online-Plattformen Berichte etwa über die Entfernung von etwaigen Netzwerken, die verdeckte Einflussnahme ausgeübt haben.
- **Fact-Checking Initiativen:** Einige Anbieter von Onlinediensten führen eigeninitiativ und in Kooperation mit einem Netzwerk an Partnern Faktenchecks durch. Inhalte werden mit Warnhinweisen versehen und die Verbreitung im Feed stark reduziert, sodass die Wahrscheinlichkeit geringer ist, dass Personen diese auffinden. Wenn ein Beitrag mit einem Faktenprüf-Label versehen ist, klicken 95 Prozent der Leute nicht weiter, um ihn anzusehen. Diese Maßnahmen regen Nutzerinnen und Nutzer zum bewussteren und kritischen Umgang mit Inhalten an. Einige Unternehmen bieten auch Programme zur finanziellen Förderung von Fact-Checking Organisationen an.

  Dienste kollaborieren mit Fact-Checking Organisation aus dem Bereich der Zivilgesellschaft, den Medien und dem Journalismus, um Desinformation gezielt zu bekämpfen. Hierbei stellt die Plattform Correctiv und die Agenturen dpa oder AFP zentrale Akteure dar mit denen Mitgliedsunternehmen kooperieren. Diese sind oft direkt in den Moderationsprozess von Inhalten eingebunden und unterstützen dabei, Desinformation zu identifizieren und schädliche Inhalte zu entfernen. Ein Beispiel bietet das European Fact-Checking Standards Network (EFCSN), dass vor den EU-Wahlen ein Projekt mit Unterstützung von Meta zur Identifizierung und Entlarvung von KI-generierten und digital veränderten Inhalten startete.
- **Entwicklung neuer Lösungen:** Mitgliedsunternehmen sind in die Entwicklung neuer Konzepte und technischer Lösungen der Desinformationsbekämpfung involviert. Im Bereich der Mitgliedschaft des Bitkom engagieren sich Unternehmen mit Blick auf die Entwicklung technischer Lösungen und nicht technischer Maßnahmen in

Kooperation mit Forschungseinrichtungen, Universitäten und zivilgesellschaftlichen Organisationen. Mit dem Innovationsprogramm für nachhaltige Lösungen, »X-Creation«, arbeitet unter anderem die Deutsche Telekom an innovativen Lösungen. Dieses Programm basiert auf einem Co-Creation-Ansatz mit zivilgesellschaftlichen Partnern und beinhaltet verschiedene Schritte, von der Problemanalyse bis zur Prototypenentwicklung. Unter dem Projekt »Schutz unserer Demokratie vor den Risiken KI-gestützter Desinformation« wurden dieses Jahr erste Konzepte für eine technische Lösung entwickelt. Die Human Centered Technology Community der Deutschen Telekom organisiert darüber hinaus Barcamps, um den Austausch zwischen Menschen aus verschiedenen Bereichen zu fördern und Denkanstöße für zukünftige Lösungsprojekte zu geben.

- **Medienkompetenz, Aufklärung und Pre-Bunking:** Die Stärkung einzelner Individuen hinsichtlich ihrer Medienkompetenz ist ein zentraler Hebel bei der Bekämpfung von Desinformation. Medienkompetenz meint »Wissen über Medien und ihre Funktionsweisen sowie ein kompetentes, selbstbestimmtes Handeln mit Medien«[14]. Hierbei sind öffentliche Bildungseinrichtungen gefragt, ein systematisches Angebot zum Erwerb von Medienkompetenz und zum Umgang mit Nachrichten anzubieten.
  - Aber auch Unternehmen fördern Medienkompetenz zum Beispiel durch Werbekampagnen oder Hinweise in den Diensten, die auf Aufklärung und Sensibilisierung ausgelegt sind und den kompetenten Umgang mit Informationen befördern sollen. So fördert zum Beispiel die Initiative »Teachtoday« der Deutschen Telekom die sichere und kompetente Mediennutzung und klärt darin auch über Desinformationen auf. Sie gibt Eltern und Lehrenden Material an die Hand, um Kinder und Jugendliche digital stärker zu machen. Das zu »Teachtoday« gehörende digitale Medienmagazin »Scroller« richtet sich direkt an Kinder und bietet ein interaktives, spielerisches Interface, um Medienkompetenz, auch mit speziellem Bezug zu Deepfakes, zu fördern. Mit ihrer Initiative #GegenHassImNetz und dazugehörigen Kampagnenspots und Themenspecials auf der Website, sensibilisiert die Deutsche Telekom bereits seit 2020 zu den gesamtgesellschaftlichen Herausforderungen von Hassrede und Desinformation.
  - Die Vodafone Stiftung engagiert sich mit der Initiative »Klick« seit 2020 für die Sensibilisierung von jungen Menschen im Umgang mit Desinformation. Dabei arbeitet die Initiative mit bekannten Influencerinnen, um das Bewusstsein für Desinformation gezielt in die Communitys junger Menschen zu tragen. Auch die von der Vodafone Stiftung geförderte Bildungsinitiative »Coding for Tomorrow« bietet Workshops und Materialien für Lehrkräfte an, mit denen junge Menschen für Desinformation sensibilisiert werden und lernen, wie sie diese erkennen und mit den Falschmeldungen umgehen.
  - Mit der Initiative »WAKE UP – gemeinsam gegen Desinformation« stärkt O2 Telefónica die Medien- und Nachrichtenkompetenz von Kindern, Jugendlichen, Lehrkräften und Eltern durch digitale Bildungsangebote wie Webvideos,

[14] Narr, Kristin et al., Medienkompetenz und Digital Literacy, 22.02.2021, Medienkompetenz und Digital Literacy | Politische Bildung in einer digitalen Welt | bpb.de.

edustories und digitale Tafelbilder sowie durch Workshops. Im Rahmen des »Jugenddialogs« fördert die Initiative den Austausch junger Menschen mit politischen Entscheidungsträgern und Entscheidungsträgerinnen zu Themen wie politischer Kommunikation, Desinformation und den Auswirkungen von Social Media auf die Meinungsbildung. Zusätzlich unterstützt O2 Telefónica mit »Digital mobil im Alter« gezielt ältere Menschen im Umgang mit Desinformation. Das Informationspaket »Faktisch betrachtet – Fit gegen Fake News« umfasst Erklärvideos, Checklisten zur Identifikation von Falschinformationen, das interaktive Online-Quiz »Dem Fake auf der Spur!« und Gesprächsreihen mit Vertreterinnen und Vertretern aus Politik, Medien, Wissenschaft und Bildungseinrichtungen zu Themen wie Desinformation und KI, um die Medienkompetenz zu fördern.

  - Im Vorfeld der Europawahlen 2024 startete Google und dessen Tochter Jigsaw eine Videokampagne nach der »Prebunking«-Methode. Die Videos mit dem Claim »Lass Dich nicht manipulieren« wurden im Mai und Juni fünf Wochen lang auf YouTube, Instagram und Facebook gezeigt. Sie klärten über häufige Techniken der Manipulation auf: Dekontextualisierung, Rufschädigung sowie die »Sündenbock-Methode«.

  - Im Rahmen einer laufenden Kampagne in Zusammenarbeit mit der Online-Journalistenschule Reporterfabrik, informiert TikTok Nutzerinnen und Nutzer darüber, wie sie Desinformation über den Russland-Ukraine-Krieg rechtzeitig erkennen und wie sie dazu beitragen können, die Verbreitung von Falschinformationen zu verhindern.

- **Einsatz von qualifizierten Webseitenzertifikaten (QWAC):** QWACs sind EV-ähnliche Webseitenzertifikate und ein Vertrauensdienst gemäß der eIDAS-Verordnung (VO (EU) 910/2014). Sie machen die Verantwortlichen von Online-Angeboten eindeutig identifizierbar – nicht zuletzt, weil ihrer Ausstellung umfangreiche Identitätsprüfungen vorausgehen. Wurde für eine Website ein QWAC ausgestellt, können Internetnutzende sicher sein: Hinter der Seite steht eine identifizierte und hierdurch grundsätzlich vertrauenswürdige Organisation. Zudem bleiben die Daten bei der Eingabe durch Verschlüsselung geschützt. Durch die in diesem Jahr in Kraft getretene Novellierung der eIDAS-Verordnung werden Browser verpflichtet, zeitnah QWAC deutlich und verbraucherfreundlich anzuzeigen, beispielsweise durch einen grünen Haken im Suchfenster. Wenn Internetnutzende eine Webseite durch einen Link aufrufen, wird bei Verwendung eines QWAC die validierte Identität des Domaininhabers und nicht mehr nur die Domain angezeigt, welche bei Täuschungsabsicht leicht verwechselt werden kann. So können Nutzende auf den ersten Blick erkennen, ob, sie auf der echten Webseite sind und ob es sich um eine authentische Informationsquelle handelt oder um eine dubiose, unsichere Verbindung.

II. Deepfakes

Auf die praktischen Herausforderungen, die von Deepfakes ausgehen, wird dieses Kapitel gesondert eingehen, denn Deepfakes können ein technisches Mittel der Desinformationsgenerierung darstellen. Hieraus ergeben sich andere Voraussetzungen zur rechtlichen Handhabung und Ergreifung geeigneter Gegenmaßnahmen als im Verhältnis zu herkömmlicher Desinformation. Maßnahmen gegen Deepfakes umfassen bereits heute weitreichende technische Methoden.

1. Definition und Abgrenzung

Durch generative KI entstehen grundsätzlich große Potenziale, vor allem im Bereich der Content-Erstellung und z. B. bei der Bildbearbeitung oder Visualisierung von Inhalten und Diensten. Ihr Einsatz dient der künstlerischen Bearbeitung. Diese Potenziale können aber auch missbraucht werden, z. B. indem Bild- oder Audiomaterial ohne Einverständnis weiterverwendet wird und es hierdurch zu Persönlichkeitsrechtsverletzungen kommt. In diesem Zusammenhang können Deepfakes eine technische Methode darstellen, um Desinformationen zu erzeugen.

Die europäische KI-Verordnung (AI Act) definiert in Art. 3 Abs. 60 Deepfakes als »einen durch KI erzeugten oder manipulierten Bild-, Ton- oder Videoinhalt, der wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähnelt und einer Person fälschlicherweise als echt oder wahrheitsgemäß erscheinen würde«. Der Begriff Deepfake bezieht sich dabei auf eine Kombination aus Deep Learning und Fake.

Beim Deepfake unterscheidet man die folgenden Arten:

1. Face Swapping: Manipulation von Bildern und Videos
2. Body Pupperty: Manipulation von Bewegungen
3. Voice Swapping: Manipulieren von Audioinhalten

Regelmäßig werden die einzelnen Arten miteinander kombiniert, sodass die Erstellung von Deepfakes z. B. zwei Prozesse umfasst: Gesichtstausch und das Klonen der Stimme. Mithilfe technologischer Verfahren erstellte Bild-, Audio- oder auch Videodarstellungen sind keineswegs neuartige Erscheinungen und existieren bereits seit spätestens der Einführung von Bild-/Audio- oder Videobearbeitungssoftware. Die mit den modernen, oft KI-gestützten Technologien einhergehende Vereinfachung solcher Manipulationen macht es aber notwendig, auf die Potenziale und Herausforderungen erneut hinzuweisen.

In der Diskussion rund um Deepfakes und potenzielle Regulierungen gilt es zu unterscheiden zwischen »einfachen« Deepfakes, welche zur (unschädlichen) Unterhaltung aber auch der Effizienzsteigerung oder Unterstützung in Arbeitsprozessen dienen, und solchen, welche negative Auswirkungen auf den demokratischen Diskurs, gesellschaftlichen Zusammenhalt, wirtschaftliche Stabilität oder die Rechte (anderer) Betroffener haben können. Deepfakes sind nicht per se gefährlich oder demokratiegefährdend, können es aber sein und in diesem Zusammenhang oder in weiteren, vorgenannten Konstellationen missbraucht werden. Hierbei ist erneut der Hinweis von zentraler Bedeutung, dass Deepfakes nicht nur im

politischen, sondern auch im wirtschaftlichen Kontext vermehrt genutzt werden, um einen systematischen Schaden herbeizuführen. Bevor der sodann folgende Abschnitt auf technische und nicht technische Maßnahmen der Deepfake-Bekämpfung genauer eingeht, sollen noch einmal gesondert die politischen und wirtschaftlichen Herausforderungen im aktuellen Zeitgeschehen dargestellt werden.

2. Neue Herausforderung – Deepfakes

Diese durch Künstliche Intelligenz generierten oder manipulierten Inhalte, die Personen, Objekte oder Ereignisse täuschend echt darstellen, eröffnen neue Dimensionen der Desinformation. Ursprünglich in harmlosen Kontexten und mit unschädlichen Effekten wie in der Unterhaltung eingesetzt, werden Deepfakes auch in politischen, gesellschaftlichen und wirtschaftlichen Kontexten genutzt, um Desinformation zu verbreiten und öffentlichen oder individuellen persönlichen Schaden anzurichten.

Die Technologie hinter Deepfakes hat sich rapide entwickelt, sie ist zugänglicher geworden und ihre Erkennung schwieriger. Das kann besorgniserregende Auswirkungen auf den demokratischen Diskurs haben und im Übrigen, da gefälschte Inhalte »täuschend echt« wirken, schnell Verbreitung finden und das Vertrauen in veröffentlichte Informationen untergraben. In politisch-gesellschaftlich herausfordernden, oft emotional erheblich aufgeladenen Zeiten könnten solche Technologien genutzt werden, um Wahlen zu beeinflussen oder gesellschaftliche Spaltungen zu vertiefen.

Angesichts dieser Herausforderungen ist es entscheidend, rechtliche und technologische Gegenmaßnahmen zu verstärken. Im nächsten Abschnitt werden spezifische Maßnahmen gegen die Generierung, Herstellung und Verbreitung von Deepfakes diskutiert. Dies umfasst sowohl legislative Ansätze als auch innovative Technologien zur Erkennung und Bekämpfung dieser Manipulationen, um sicherzustellen, dass die Integrität öffentlicher Diskurse weitestgehend geschützt wird.

3. Maßnahmen gegen Deepfakes

3.1 Rechtliche Handhabung und Regulierung

Da Deepfakes dazu verwendet werden können, Desinformation zu erzeugen, gilt die Einschätzung zur rechtlichen Handhabung zum Thema Desinformation entsprechend auch für Deepfakes. Unter der Prämisse, dass Deepfakes KI-generiert sind, gelten für Deepfakes darüber hinaus die Transparenzpflichten aus dem AI Act als KI-spezifisches Regularium. Bei Deepfakes lohnt darüber hinaus auch eine spezielle Betrachtung des Datenschutz- und Persönlichkeitsrechts.

AI Act: Transparenzvorgaben für KI-generierte Inhalte

Nach Art. 50 Abs. 4 AI Act müssen Betreiber von KI-Systemen, die Inhalte erzeugen oder manipulieren, die Deepfakes sind, offenlegen, dass die Inhalte künstlich erzeugt oder

manipuliert wurden. Die Pflicht zur Offenlegung haben auch Betreiber von KI-Systemen, die Texte erzeugen oder manipulieren, die veröffentlicht werden, »um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren«. Diese Transparenzpflicht muss den betroffenen natürlichen Personen »spätestens zum Zeitpunkt der ersten Interaktion oder Aussetzung in klarer und eindeutiger Weise bereitgestellt werden (Art. 50 Abs. 5 AI Act). Es lässt sich also festhalten, dass der EU-Gesetzgeber für Deepfakes Kennzeichnungspflichten vorsieht, mit welchen den Täuschungsrisiken durch den technischen Fortschritt begegnet werden soll. Weitere Regelungen betreffend Deepfakes sieht der AI Act nicht vor. Sie werden nicht als Hochrisikoanwendungsfall eingestuft.

Persönlichkeits- und Datenschutzrecht

Das allgemeine Persönlichkeitsrecht schützt u. a. »vor der Verbreitung eines technisch manipulierten Bildes, das den Anschein erweckt, ein authentisches Abbild einer Person zu sein.«[15] Gleiches dürfte auch für das Recht am eigenen Wort gelten. Zwar hat die betroffene Person kein Recht darauf, nur so von anderen wahrgenommen zu werden, wie sie sich das wünscht, dennoch kann ihr aus dem allgemeinen Persönlichkeitsrecht u.U. ein Unterlassungsanspruch gegen die Verbreitung des Deepfakes zustehen. Allerdings kann wiederum in der Meinungs-, Presse- oder Kunstfreiheit ein Rechtfertigungsgrund für das Verwenden des Deepfakes liegen (etwa bei erkennbar satirischen Darstellungen).

Die Herstellung und Verbreitung von Deepfakes können datenschutzrechtlich eine Verarbeitung personenbezogener Daten darstellen. Für eine solche Verarbeitung muss in aller Regel eine Rechtsgrundlage nach Art. 6 DSGVO vorliegen. Infrage kommen einerseits die Einwilligung der betroffenen Person nach Art. 6 (1) Buchst. a DSGVO oder z. B. im Rahmen von der Erfüllung eines Vertrages nach Art. 6 (1) Buchst. B DSGVO sowie das Vorliegen einschlägiger gesetzlicher Erlaubnistatbestände. Im Rahmen von Deepfakes ist im Zusammenhang mit parodistischen bzw. satirischen Deepfakes speziell die Privilegierung in Art. 85 DSGVO zu nennen. Besonders im Bereich von Rundfunk und Presse wurden entsprechende Privilegierungen (und damit Erlaubnistatbestände) geschaffen. Darunter sind mitunter die Regelungen in §§ 12 und 23 MStV zu nennen.[16] Bis auf die Regelungen im DSA und im AI Act finden sich also für KI-generierte Deepfakes keine spezifischen Regeln, die nicht schon aus dem Kontext von Inhaltsmanipulation (ohne KI-Technologie) bekannt sind.

Strafrechtliche Bewertung

Einen großen Anstoß in der Bekämpfung von missbräuchlichen Deepfakes hat nun der Freistaat Bayern geleistet. Am 14. Mai 2024 hat der Freistaat einen Gesetzesantrag im Bundesrat eingereicht, der die strafrechtliche Ahndung von Persönlichkeitsrechtsverletzungen durch Deepfakes zum Ziel hat. So soll ein neuer § 201b StGB eingeführt werden, der Freiheitsstrafen von bis zu zwei Jahren oder

[15] BVerfG NJW 2005, 3271, 3272.
[16] Zudem haben auch einzelne Länder in entsprechenden Landesdatenschutzgesetzen Erlaubnistatbestände hinsichtlich der Personendatenverarbeitung zu u.a. journalistischen Zwecken vorgesehen (z.B. Art. 38 BayDSG).

Geldstrafen für die Erstellung und Verbreitung von Deepfakes vorsieht. In schwerwiegenden Fällen, wie etwa der Verbreitung pornografischer Deepfakes, kann die Strafe auf bis zu fünf Jahre Freiheitsstrafe erhöht werden. Ausnahmen sollen für Deepfakes gelten, die in Wahrnehmung überwiegender berechtigter Interessen erstellt werden, wie etwa im Rahmen von Kunst, Forschung, Lehre oder Berichterstattung. Eine vergleichbare Strafbarkeitsregelung für Deepfakes gibt es bislang noch nicht.

3.2 Maßnahmen zur Bekämpfung von Deepfakes

Maßnahmen, die Unternehmen gegen Deepfakes ergreifen, umfassen technische sowie nicht technische Maßnahmen, die in verschiedene Phasen der Deepfake-Generierung, Verbreitung und Erkennung zum Einsatz kommen. Im Verhältnis zu Desinformation bieten Deepfakes bereits mehr technische, automatisierte Ansatzmöglichkeiten, die der Entstehung und Verbreitung entgegenwirken können.

- **Reduzierung der Quellen und Reichweite für Deepfake-Daten:** Als präventive Maßnahmen gilt die Einschränkung der Verfügbarkeit und der konsequente Schutz von Daten, sodass die unerlaubte Nutzung von Daten zur Generierung von Deepfakes eingedämmt wird. Dies umfasst z. B. öffentlich zugängliche Daten von Social-Media-Plattformen, Websites und Nachrichtenagenturen, personenbezogene Nutzerdaten und synthetische Daten.
- **Entfernung von richtlinienwidrigen Inhalten:** Die Richtlinien und Nutzungsbedingungen von Online-Diensten umfassen weitreichende Regelungen hinsichtlich des Umgangs mit Deepfakes und manipulierten Inhalten. Rechtswidrige Inhalte werden demnach entfernt, sofern die rechtlichen Voraussetzungen vorliegen.
- **Labeling von KI-generierten Inhalten:** Onlinedienste bieten Inhalte-Erstellern Funktionen an, um KI-generierte Inhalte entsprechend zu kennzeichnen. Dies erhöht die Transparenz im Umgang mit Inhalten auf Online-Plattformen und auf Suchmaschinen und ermöglicht einen differenzierten Umgang mit Inhalten. Diensteanbieter erfüllen damit ihre Pflichten aus dem AI Act, geben Influencerinnen und Influencern aber auch unterstützende Tools an die Hand, um diese Pflichten effektiv wahrzunehmen.
- **KI-gestützte Erkennung von Deepfakes:** Auch KI-gestützte Tools spielen eine wichtige Rolle bei der Erkennung und Vermeidung von Deepfakes. Diese Tools analysieren Inhalte auf charakteristische Anzeichen von Manipulation, die für das menschliche Auge und Ohr oft nicht erkennbar sind. Durch den Einsatz von maschinellem Lernen und digitalen neuronalen Netzen können solche Systeme Deepfakes mit hoher Genauigkeit identifizieren und zur Verhinderung ihrer Verbreitung beitragen.
- **Unterstützende Funktionen:** Darüber hinaus gibt es zahlreiche Funktionen, die Verbrauchern helfen, authentische Inhalte von nicht authentischen Inhalten besser zu unterscheiden.
- **Einsatz von Content-Credentials:** Durch das Auslesen von Content-Credentials bzw. den Metadaten von Bild- und Audiodateien kann die Quelle von Inhalten effektiv identifiziert und KI-generierte Inhalte können automatisch gekennzeichnet werden.

Diese Lösungen sind aktuell am vielversprechendsten. Das »Blocken« bzw. Sperren der Verwendung bestimmter Daten zur Verarbeitung und Erzeugung von Deepfakes ist bisher noch nicht möglich.

- Adobe hat 2019 gemeinsam mit zahlreichen Partnern die Content Authenticity Initiative (CAI) ins Leben gerufen, welche sich unter anderem der Einführung und Umsetzung der Content-Credentials-Technologie auf Basis eines offenen Standards widmet. Neben großen Anbietern von Software und Online-Diensten umfasst die Initiative auch zivilgesellschaftliche Akteure und Nachrichtenverlage. Es handelt sich dabei um einen Zusammenschluss von Medien- und Technologieunternehmen, die zusammenarbeiten, um einen offenen Industriestandard rund um die Authentizität und Rückverfolgbarkeit von Bild-, Video- und Audiodateien zu etablieren. Durch die Integration digitaler Herkunftsnachweise kann Transparenz sichergestellt werden. Nutzerinnen und Nutzer können dadurch informierte Entscheidungen über die Authentizität von digitalen Inhalten auf Basis von verifizierbaren Informationen treffen.

- In diesem Zusammenhang kooperieren Diensteanbieter innerhalb der Coalition for Content Provenance and Authenticity (C2PA). C2PA hat mit Projektpartnern wie Adobe, Microsoft, Google, Meta, Intel etc. ein Verifizierungstool (Open Source) entwickelt. Dieses bietet einen offenen technischen Standard, der es Verlegern, Urhebern und Verbrauchern ermöglicht, die Herkunft verschiedener Medientypen zurückzuverfolgen.[17] Nachdem TikTok bereits die Content-Credentials-Technologie für KI-generierte Inhalte einsetzt, hat vor Kurzem auch LinkedIn Content-Credentials implementiert. Auch Google hat angekündigt, Content Credentials in einige Produkte zu integrieren. Das ist das erste Mal, dass Content-Credentials, und damit Informationen über die Entstehung und Verarbeitung von Inhalten, für Endnutzerinnen und -nutzer einsehbar sind.

- Die Bundesdruckerei-Tochter D-Trust entwickelt darüber hinaus Gerätezertifikate für Kameras, auf deren Basis die Authentizität von Mediendateien ab ihrer Aufnahme durch digitale Signaturen gesichert wird. Für den Einsatz von Content-Credentials werden die Gerätezertifikate für jede Kamera individuell erzeugt und in dieser auf einem sicheren Chip abgelegt. Bereits beim Erfassen eines Bildes, Videos oder einer Tonaufnahme versehen die Zertifikate den entsprechenden Inhalt und den angehefteten Satz von Metadaten wie Urheberschaft, Ort, Datum oder Uhrzeit mit einer digitalen Signatur.

- **Weitere Maßnahmen zur Überprüfung der Authentizität von Content:** Insbesondere im Bereich der Authentifizierung von Bild- und Audiomaterial werden derzeit weitreichende technische Lösungen im Bereich der Deepfake-Bekämpfung entwickelt.

  - Die Nürnberger Firma BioID GmbH entwickelt Software zur Erkennung von Deepfakes in Foto- und Videomaterial. In Echtzeit wird analysiert, ob es sich um echtes, nichtmanipuliertes Material oder um KI-generierte/veränderte Inhalte

[17] Coalition for Content Provenance and Authority: An open technical standard providing publishers, creators, and consumers the ability to trace the origin of different types of media, Overview - C2PA.

handelt. Diese Software wird als weltweiter, DSGVO-konformer Service angeboten und wurde in einem vom BMBF geförderten Forschungsprojekt erarbeitet.

  - Darüber hinaus hat die Bundesdruckerei GmbH im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten Forschungsprojektes Fake-ID mit Partnern spezielle Detektoren entwickelt, um falsche und manipulierte Identitäten einfacher erkennen zu können. Der entstandene Demonstrator in Form einer Softwareplattform analysiert zu welchem Grad Bewegtbildmaterial authentisch ist. Mithilfe linearer sowie KI-Verfahren werden Bild- und Videodatenströme zur Laufzeit mithilfe trainierter Algorithmen auf verschiedene Echtheitsmerkmale hin analysiert. Im Rahmen eines weiteren Forschungsprojektes hat die Bundesdruckerei gemeinsam mit Konsortialpartnern die Sicherheit und Vertrauenswürdigkeit von mobilen und eingebetteten KI-Systemen analysiert und einen Prototyp zur Echtzeiterkennung von Deepfake-Angriffen in Videokonferenzen entwickelt. Dieser nutzt visuelle Selbsterkennung als biometrischen Identifikationsmechanismus.

- **Wasserzeichen und digitale Signaturen:** Eine weitere technische Maßnahme besteht in der Implementierung von digitalen Wasserzeichen und Signaturen. Diese Technologien ermöglichen es, in digitale Inhalte unsichtbare Markierungen einzufügen, die nachträglich überprüft werden können, um Manipulationen zu erkennen. Solche Wasserzeichen können fälschungssicher gestaltet und schwer entfernbar gemacht werden, ohne die Qualität der Inhalte zu beeinträchtigen.
  - Meta verwendet digitale Wasserzeichen und IPTC-Metadaten, um Bilder aus Benutzertools wie Imagine zu markieren, die das Potenzial haben, realistische, KI-generierte oder bearbeitete Bilder zu erstellen.
  - SynthID, ein Beta-Tool von Google DeepMind, bettet ein digitales Wasserzeichen direkt in KI-generierte Bilder und Audiodaten ein.
- **Echtzeit-Überwachung und Analyse:** Die Entwicklung von KI-gestützten Systemen zur Echtzeit-Überwachung und Analyse von Inhalten stellt eine weitere wirksame Maßnahme dar. Solche Systeme können verdächtige Inhalte, die möglicherweise Deepfakes sind, in Echtzeit erkennen und markieren. Diese Technologien analysieren die Merkmale von Audio-, Video- und Bilddateien und vergleichen sie mit bekannten Mustern von Deepfakes. Dadurch kann die Verbreitung manipulierter Inhalte frühzeitig erkannt und gestoppt werden. Die Entwicklung und kontinuierliche Weiterentwicklung von Werkzeugen zur Durchführung von Echtzeitanalysen stellt angesichts des raschen Fortschritts der Deepfake-Technologie jedoch eine fortwährende Herausforderung dar.[18]
- **Blockchain-Technologie zur Rückverfolgbarkeit:** Die Blockchain-Technologie bietet ebenfalls vielversprechende Ansätze zur Verhinderung der Verbreitung von Deepfakes. Durch die dezentrale Speicherung von Informationen zur Herkunft und Authentizität digitaler Inhalte in einer Blockchain kann die Integrität dieser Daten

[18] 4 Wege zum Zukunftsschutz gegen Deepfakes im Jahr 2024 und darüber hinaus, Feb. 2024: https://www.weforum.org/agenda/2024/02/4-ways-to-future-proof-against-deepfakes-in-2024-and-beyond/.

sichergestellt werden. Jede Veränderung eines Inhalts würde sofort auffallen, da sie in der Blockchain vermerkt wird. Diese Methode ermöglicht eine transparente und fälschungssichere Nachverfolgung der Geschichte digitaler Inhalte.

- **Stärkung von Medienkompetenz und Bildungskampagnen:** Es ist entscheidend, dass der kritische Umgang mit Informationen im Netz weiter gestärkt wird, dies gilt auch für den Umgang mit Deepfakes. Hierzu starten Anbieter von Onlinediensten in Kooperation mit zivilgesellschaftlichen Organisationen öffentlichkeitswirksame Kampagnen, um das Bewusstsein für potenziell irreführende und nachweislich falsche Inhalte und neuen Herausforderungen rund um Deepfakes zu schärfen:
  - Im Kontext des Superwahljahres 2024 hat sich zum Beispiel Microsoft durch die Initiative »Elections & Deepfakes: Public Awareness Campaign: Check. Recheck. Vote« gegen den Einfluss von Deepfakes auf Wahlen stark gemacht. Die Initiative verfolgt das Ziel, über das Wahlrecht und Möglichkeiten der Einflussnahme durch Deepfakes zu sensibilisieren.[19]
  - Darüber hinaus bietet Microsoft Möglichkeiten der Schulung von politischen Stakeholdern und Kampagnenmitarbeiterinnen und -mitarbeitern über den Umgang mit Deepfakes im Kontext von Wahlen an. Über technische Unterstützungstools lässt sich zum Beispiel eine sichere Inhalte-Verwaltung und Mediendatenbank erstellen.
  - Die Initiative »Digital Mobil im Alter« von o2 Telefónica sensibilisiert ältere Menschen für den Umgang mit Deepfakes durch digitale Erlebnisparcours und Workshops zu verschiedenen Manipulationstechniken.
- **Mitarbeitersensibilisierung und Aufklärungsarbeit im Unternehmen:** Unternehmen bieten vermehrt interne und externe Schulungen zum Thema Deepfakes und Desinformation an, um mehr Bewusstsein für das Thema zu schaffen. Für viele Unternehmen wie auch BioID steht hierbei die Aufklärung über den Deepfake-Betrug im Vordergrund. Da mittlerweile vermehrt Unternehmen von Deepfake-Angriffen betroffen sind, müssen Mitarbeiterinnen und Mitarbeiter zu Cybersicherheitsrisiken stärker geschult werden.
  - Darüber hinaus setzen sich Unternehmen mit der Nutzung von KI verstärkt auseinander, indem sie unternehmensinterne KI-Leitlinien verfassen. Diese gewährleisten einen sicheren Umgang mit KI-Systemen. Die Resilienzstärkung übernimmt zum Beispiel die Deutsche Telekom durch sogenannte »Lex Sessions: Lernen von Experten«. Auch die Awareness Akademie der Telekom Security (T-Sec) sensibilisiert Mitarbeitende zu Themen wie Deepfake-unterstütztem CEO-Fraud. Die T-Sec bietet das »Security Activity Book AwareNessi« für Kinder und Erwachsene an, das spielerisch zu digitalen Themen, darunter auch Desinformationen und Deepfakes, aufklärt und digitale Kompetenzen stärkt.

- **Partnerschaften und Kooperationen in der digitalen Wirtschaft:**
  - AI Elections Accord 2024: 20 führende Technologieunternehmen, darunter Adobe, Amazon, Google, IBM, Meta, Microsoft, OpenAI, TikTok etc. verpflichten sich zur Zusammenarbeit bei der Erkennung und Bekämpfung schädlicher KI-Inhalte. Es handelt sich dabei um eine weitreichende Partnerschaft von Technologieunternehmen und Plattformen, die die Bekämpfung des missbräuchlichen Einsatzes von KI im Superwahljahr 2024 weiter voranbringen wollen. Besonderer Fokus der Initiative sind daher KI-generierte Audio-, Video- und Bildinhalte, die das Aussehen, die Stimme oder die Handlungen von politischen Kandidaten, Wahlhelfern, Wahlbeamten und anderen wichtigen Akteuren einer demokratischen Wahl vortäuschen oder verfälschen.

    Wichtigste Maßnahmen der Initiative:

    - Entwicklung und Umsetzung von Technologien zur Minderung der Risiken im Zusammenhang mit schädlichen KI-generierten Wahlinhalten und Entwicklung von Open-Source-Lösungen
    - Identifikation der Authentizität und der Herkunft von Inhalten wie Bild- und Audiomaterial (mithilfe von Verfahren wie C2PA und digitalen Wasserzeichen)
    - Identifikation von Risiken durch datengetriebene Modelle und Ausbau des Verständnisses über schädliche Inhalte wie Deepfakes
    - Bemühung um einen angemessenen Umgang mit diesen Inhalten und Anwendung von Erkennungstechnologien
    - Förderung der branchenübergreifenden Resilienz gegenüber betrügerischen KI-Wahlinhalten durch das Teilen von Best Practices (cross-industry-resillience)
    - Transparenz gegenüber der Öffentlichkeit hinsichtlich des Umgangs des Unternehmens mit diesen Inhalten
    - Fortsetzung der Zusammenarbeit mit einer Vielzahl von Organisationen der globalen Zivilgesellschaft und Wissenschaftlern
    - Unterstützung von Bemühungen zur Förderung des öffentlichen Bewusstseins, der Medienkompetenz und der gesamtgesellschaftlichen Resilienz

III. Bayern-Allianz gegen Desinformation

1. Desinformation schädigt Demokratie

Bayern hat sich dem Kampf gegen Desinformation im Netz verschrieben, weil sich der Informationsfluss vom Analogen zunehmend ins Digitale verlagert. Demokratische Institutionen und ihre Vertreterinnen und Vertreter müssen dort Präsenz zeigen, wo Desinformation entsteht und mit enormen Reichweiten nahezu mühelos verbreitet wird.

Die bayerische Staatsregierung setzt sich dafür ein, dass das Internet nicht zum rechtsfreien Raum wird, in dem politische Geschäftemacher durch Fake News in sozialen Medien die Oberhand über den demokratischen Diskurs gewinnen. Gerade im Umfeld von Wahlen ist ein Anstieg von gezielter Desinformation zu beobachten. Umso mehr gilt es, die Integrität und Fairness demokratischer Prozesse auch im digitalen Raum zu gewährleisten.

2. Der bayerische Ansatz: Mit konkreten Aktionen gegen Fake News

Ausgangspunkt der Bayern-Allianz gegen Desinformation war das Rekordwahljahr 2024. Mit Blick auf die Europawahl hatte die Europäische Kommission eine übergreifende Strategie verabschiedet. Zahlreiche Tech-Unternehmen haben sich auf der Münchner Sicherheitskonferenz 2024 dem Munich Tech Accord verpflichtet. Die bayerische Staatsregierung knüpft an diese Aktionen mit konkreten Maßnahmen an, woran die Partner aus Tech-Wirtschaft, Zivilgesellschaft, Politik, Wissenschaft und Medien zur Umsetzung beitragen.

Der bayerische Staatsminister für Digitales Dr. Fabian Mehring, MdL initiierte zusammen mit dem bayerischen Staatsminister des Innern, für Sport und Integration Joachim Herrmann, MdL die Bayern-Allianz gegen Desinformation. Diese Initiative, am 8. Mai mit Vorlauf zu den Europawahlen gestartet, stützt sich zunächst auf drei Säulen:

- Ein Bündnis mit internationalen Plattformbetreibern und Tech-Unternehmen
- Ein umfangreiches Paket staatlicher Maßnahmen
- Eine Kooperation mit etablierten Medien

Derzeit wird die Bayern-Allianz gegen Desinformation um zwei neue Säulen erweitert: eine Verankerung in die Breite der bayerischen Zivilgesellschaft sowie die Assoziierung politischer Organisationen. Sie will weitere Partner, insbesondere aus Wirtschaft, Wissenschaft und Zivilgesellschaft gewinnen, um Konzepte für wirksame Gegenmaßnahmen voranzutreiben und ihre organisatorische Struktur auszubauen. Danach sollen im Vorfeld von Bundestagswahl (Herbst 2025) und bayerischer

Kommunalwahl (Frühjahr 2026) die breite Ausrollung von Maßnahmen gegen Desinformation folgen. Gemeinsam werden sich die Akteure mit Angeboten gegen Desinformation engagieren, unter anderem um breite Bevölkerungsteile, beispielsweise über Vereins- und Ehrenamtsstrukturen, auf die Gefahren von Falschnachrichten und gezielter Manipulation im Netz zu sensibilisieren.

Die Bayern-Allianz gegen Desinformation verfolgt zwei Ziele:

- Sensibilisierung von Bürgerinnen und Bürgern für das Bedrohungsphänomen der Desinformation. Mit gezielten Maßnahmen soll die Resilienz der Bevölkerung erhöht werden, damit Desinformationskampagnen erst gar nicht verfangen.
- Einsatz von technologischen Maßnahmen zur Sichtbarmachung und Reduktion von Desinformation im Netz.

3. Die Tech-Unternehmen: Starke Partner im Zentrum

Die als Gründungspartner beteiligten sieben Tech-Unternehmen engagieren sich in der Bayern-Allianz mit einer großen Bandbreite an Themen. Weitere Unternehmen werden sukzessive dazustoßen und das Spektrum erweitern und vertiefen. Zu den konkreten Maßnahmen zählen unter anderem:

- **Adobe** trägt mit dem Einsatz von Content Credentials bei: Es handelt sich um eine kostenlose Open-Source-Technologie, die wie ein »Nährwertkennzeichen« für digitale Inhalte fungiert.
- **Google** sensibilisierte im Vorfeld der Europawahl die Bevölkerung mit einer Informationskampagne nach dem »Prebunking-Konzept«: Videos mit dem Claim »Lass Dich nicht manipulieren« wurden als Social-Media-Anzeigen unter anderem in Deutschland geschaltet.
- **IBM** führt Veranstaltungen durch, um einerseits das Bewusstsein in der Bevölkerung für das Thema Desinformation und den Umgang damit zu schärfen und andererseits Fachleuten aus Bereichen wie Medien, Politik, Cybersecurity eine Plattform für den Gedanken- und Wissensaustausch zu bieten.
- **Meta** arbeitet mit unabhängigen Faktenprüfern wie Correctiv, dpa und AFP zusammen, um Desinformation zu bekämpfen.
- **Microsoft** stellte einen Werkzeugkasten bereit, um die Integrität von Inhalten sicherzustellen. Der Werkzeugkasten besteht unter anderem aus einem Melde-Formular zum Schutz von Kandidierenden vor betrügerischer KI.
- **O2 Telefónica** legt den Fokus der Maßnahmen auf ältere Menschen: digitale Spaziergänge in Bayern mit dem Themenfokus Desinformation. Spaziergang-Teilnehmerinnen und Teilnehmer versuchen Actionbound-App-Fragen u. a. zu Desinformation und Falschinformation zu lösen und werden trainiert, Informationen kritisch zu hinterfragen.
- Siemens ermöglichte Mitarbeiter-Trainings: Beispielsweise wurde an den Siemens-eigenen Berufsschulen Auszubildenden und Dual Studierende von Anfang an

sensibilisiert, vertrauenswürdige und richtige Informationsquellen (insbesondere im Zeitalter von KI-Modellen wie ChatGPT) zu identifizieren.

4. Staat und mediale Begleitung: Menschen über vielfältige Kanäle erreichen

Um möglichst viele Menschen zu erreichen, wurden noch weitere konkrete Maßnahmen unter Beteiligung einer Vielzahl von Institutionen und Organisationen durchgeführt.

Erfahrene und kompetente staatliche oder staatsnahe Einrichtungen ergänzten die Initiativen der Tech-Unternehmen: Neben den umfassenden Aktivitäten des Bayerischen Landesamtes für den Verfassungsschutz gab es Hintergrundinformationen und einen Podcast des bayerischen Staatsministeriums des Innern, für Sport und Integration. Die bayerische Landeszentrale für politische Bildungsarbeit konnte neben weiteren Maßnahmen für einen Webtalk zum Thema »Desinformation im europäischen Wahlkampf: Reale Gefahr oder mediales Schreckgespenst?« gewonnen werden. Das Staatsinstitut für Schulqualität und Bildungsforschung ISB brachte sich mit einem Themen-Portal zur politischen Bildung ein.

Der renommierte bundesweit tätige Verein Deutschland sicher im Netz e. V. erwies sich ebenfalls als sehr engagierter Partner. Dessen Expertinnen und Experten besuchten in der Vorwahlzeit verschiedene bayerische Marktplätze, um direkt mit der Bevölkerung zum Thema Fake News in Kontakt zu treten. Zudem bündelte Deutschland sicher im Netz e. V. die entsprechenden Materialien seines Digitalführerscheins, dessen Lernmodule auch eine Festigung des erworbenen Wissens mittels Prüfungsetappen ermöglichen.

Mit dem Bayerischen Rundfunk (BR), der Vereinigung bayerischer Rundfunkanbieter e. V. (VBRA) sowie dem Verband Bayerischer Zeitungsverleger e. V. (VBZV) sind auch Vertreter der bayerischen Medienlandschaft Teil der Bayern-Allianz gegen Desinformation. In Wahrung ihrer Unabhängigkeit beschränken sich diese Einrichtungen darauf, ihre Expertise mit anderen Partnern zu teilen. Beispielsweise spürt der BR24 mit dem #Faktenfuchs Desinformation und Falschbehauptungen in den sozialen Medien auf und überprüft und erklärt immer auch die Hintergründe. Mit ‚Social Listening' hat BR24 zudem das Ohr gezielt dort, wo Menschen auf Fälschungen treffen oder treffen könnten. Ziel ist es, Desinformation früh einzufangen, Menschen dafür zu sensibilisieren und so den gesellschaftlichen Schaden kleinzuhalten.

5. Ausblick für die Weiterentwicklung der Bayern-Allianz gegen Desinformation

Der Start der Bayern-Allianz vor der Europawahl war angesichts des breiten Maßnahmenspektrums und der Resonanz in der Öffentlichkeit sehr erfolgreich: Allein die Social-Media-Kampagne erreichte sechs Millionen Menschen.

Mit ungewöhnlichen Bildmotiven wie dem Schloss Neuschwanstein mitten im Urwald wurde unter dem Motto »Glaub nicht alles, was Du siehst!« zum Thema Fake News sensibilisiert.

In den kommenden Wochen und Monaten wird die Bayern-Allianz gegen Desinformation auch in Richtung der Zivilgesellschaft verbreitert: Es wird darum gehen, in den Vereinen und im Ehrenamt, in Verbänden, Kirchen, Gewerkschaften und karitativen, sozialen Gruppen neue Partner und Mitstreiterinnen und Mitstreiter zu gewinnen. Viele zivilgesellschaftliche Gruppen vor Ort in der Fläche sollen über die vorhandenen Organisationsstrukturen erreicht und sensibilisiert werden. Der »Graswurzelansatz« verfolgt das Ziel, möglichst viele Menschen in Bayern in allen Teilen des Freistaats und hinweg über alle Altersgruppen zu erreichen. Denn Desinformation ist kein Bedrohungsphänomen, welches ausschließlich eine digitalaffine Gesellschaftsgruppe betrifft. Gerade mit Blick auf die Bundestagswahl 2025 will die bayerische Staatsregierung ihren Beitrag leisten, um Desinformation und die hinter ihr stehenden, destruktiven Kräfte zu minimieren und bestmöglich die digitale Resilienz unserer Demokratie zu stärken.

Die Weiterentwicklung der Bayern-Allianz lebt vom Austausch und wissenschaftlicher Begleitung: Ein Best Practice Dialog wurde beispielsweise mit Ansprechpartnern in Österreich und Schweden angestoßen. Für die wissenschaftliche Begleitung konnte zudem das Bayerische Institut für Digitale Transformation gewonnen werden.

Im Schulterschluss mit seinen Partnern aus Wirtschaft, Gesellschaft und Wissenschaft ergreift der Freistaat Bayern wirksame Maßnahmen im Kampf gegen Desinformation und für digitale Resilienz.

Bayern-Allianz gegen Desinformation

Gerne können sich weitere Tech-Unternehmen an der Bayern-Allianz gegen Desinformation beteiligen. Aktuelle Informationen und Aufnahmemodalitäten finden Sie unter: https://www.stmd.bayern.de/themen/bayern-allianz-desinformation/

Ansprechpartner

Vera Cornette I Leiterin des Planungsstabs
E-Mail I vera.cornette@stmd.bayern.de

Dr. Rolf Bommer I Leiter des Referats Digitale Teilhabe, Digitales Leben
E-Mail I rolf.bommer@stmd.bayern.de

Bayerisches Staatsministerium für Digitales
Oskar-von-Miller-Ring 35, 80333 München

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Herausgeber
Bitkom e.V.
Albrechtstr. 10 | 10117 Berlin

Ansprechpartner/in
Luise Ritter | Referentin für Medienpolitik & Plattformen
E-Mail: l.ritter@bitkom.org

Dr. Pablo Schumacher | Bereichsleiter Digital Content & Recht
E-Mail: p.schumacher@bitkom.org

Marvin Pawelczyk | Referent Künstliche Intelligenz & Cloud
E-Mail: m.pawelczyk@bitkom.org